frontiers | Research Topics

Eye tracking for STEM education research: New perspectives

Edited by
Pascal Klein, Martin Rusek and Maike Schindler

Published in
Frontiers in Education

ISSN 1664-8714
ISBN 978-2-8325-4732-8
DOI 10.3389/978-2-8325-4732-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Eye tracking for STEM education research: New perspectives

Topic editors
Pascal Klein — University of Göttingen, Germany
Martin Rusek — Charles University, Czechia
Maike Schindler — University of Cologne, Germany

Citation
Klein, P., Rusek, M., Schindler, M., eds. (2024). *Eye tracking for STEM education research: New perspectives*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4732-8

Table of contents

frontiers | Frontiers in Education

TYPE Editorial
PUBLISHED 26 March 2024
DOI 10.3389/feduc.2024.1389962

OPEN ACCESS

EDITED AND REVIEWED BY
Lianghuo Fan,
East China Normal University, China

*CORRESPONDENCE
Martin Rusek
martin.rusek@pedf.cuni.cz

RECEIVED 22 February 2024
ACCEPTED 18 March 2024
PUBLISHED 26 March 2024

CITATION
Rusek M, Klein P and Schindler M (2024) Editorial: Eye tracking for STEM education research: new perspectives.
Front. Educ. 9:1389962.
doi: 10.3389/feduc.2024.1389962

Editorial: Eye tracking for STEM education research: new perspectives

Martin Rusek[1]*, Pascal Klein[2] and Maike Schindler[3]

[1]Department of Chemistry and Chemistry Education, Faculty of Education, Charles University, Prague, Czechia, [2]Physics Education Research, Faculty of Physics, University of Göttingen, Göttingen, Lower Saxony, Germany, [3]Special Education in Mathematics, University of Cologne, Cologne, North Rhine-Westphalia, Germany

KEYWORDS

eye-tracking (ET), STEM - science technology engineering mathematics, education research, Research Topic, innovations in research methodology

Editorial on the Research Topic
Eye tracking for STEM education research: new perspectives

The integration of eye-tracking (ET) technology into STEM education research marks a pivotal shift toward a more nuanced educational methodologies' comprehension and improvement. The Research Topic titled "Eye-Tracking for STEM Education Research: New Perspectives," introduced a series of pioneering studies that employ ET technology to dissect and understand the complex nature of learning in the realms of science, technology, engineering, and mathematics.

The issue commences with an insightful exploration of how the combination of ET and artificial intelligence is transforming the landscape of competency assessment within engineering education. The paper "Eye-Tracking and Artificial Intelligence for Competency Assessment in Engineering Education: A Review" (Ndiaye et al.) serves as a cornerstone for this edition, highlighting the interdisciplinary fusion that characterizes the subsequent contributions. This contribution also introduces a new dimension to the forefront of research in the field. Since the launch of this Research Topic, there has been a significant advancement in technology. Artificial Intelligence (AI) is undoubtedly an aspect that the global community must integrate, a notion that holds particularly true for STEM education and its research endeavors.

Further, the issue delves into the application of ET in specific STEM disciplines. In the domain of physics education, ET's role as a pivotal feedback mechanism in teacher training is explored, alongside its utility in analyzing the cognitive processes involved in understanding vector fields (Hahn and Klein). These investigations pave the way for similar explorations within chemistry education, where studies examine the relationship between pupil dilation and cognitive load during instructional video sessions (Rodemer et al.), therefore expanding the possibilities educators will soon have at their disposal to facilitate their students' problem-solving process in real time. This line of inquiry in this Research Topic was also elaborated in chemistry, showcasing ET's critical role in dissecting the problem-solving process from another perspective (Tóthová and Rusek). Additionally, the assessment of collaborative knowledge construction (Lämsä et al.) revealed various methods by which students form conceptions of scientific phenomena in their minds.

A standout contribution within mathematics education revolves around the use of ET in statistics, specifically in how students engage with data. This research sheds light on the nuanced ways students navigate statistical information, offering a fresh perspective on data interpretation and processing (Schreiter and Vogel) which is associated with students' graph interpretation processes (Thomaneck et al.) or the way they are able to link information from multiple representation (Susac et al.) – a discipline every student faces. This study also brought further confirmation into the sometimes debated (Schindler and Lilienthal, 2019) eye-mind hypothesis (Just and Carpenter, 1980).

The issue rounds off by emphasizing the significance of visual representation in STEM learning, particularly through a study on organic chemistry (Braun et al.). This research examined how students employ ET technology to navigate the drawing of complex molecular structures, underscoring the technology's value in understanding and enhancing visual learning strategies.

This Research Topic not only highlighted the multifaceted applications of eye tracking in STEM education research but also reinforces its potential to significantly enrich our comprehension of learning dynamics and instructional methods across diverse scientific disciplines. The authors provide multiple implications for further research which promises more interesting findings in the near future.

Author contributions

MR: Writing – original draft, Writing – review & editing. PK: Writing – review & editing. MS: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329.

Schindler, M., and Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educ. Stu. Mathematics* 101, 123–139. doi: 10.1007/s10649-019-9878-z

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/feduc.2022.942224

Check for updates

OPEN ACCESS

EDITED BY
Martin Rusek,
Charles University,
Czechia

REVIEWED BY
Anna Shvarts,
Utrecht University,
Netherlands
Andrew Ruis,
University of Wisconsin-Madison,
United States

*CORRESPONDENCE
Joni Lämsä
joni.lamsa@oulu.fi

SPECIALTY SECTION
This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

RECEIVED 12 May 2022
ACCEPTED 03 October 2022
PUBLISHED 20 October 2022

CITATION
Lämsä J, Kotkajuuri J, Lehtinen A, Koskinen P, Mäntylä T, Kilpeläinen J and Hämäläinen R (2022) The focus and timing of gaze matters: Investigating collaborative knowledge construction in a simulation-based environment by combined video and eye tracking.
Front. Educ. 7:942224.
doi: 10.3389/feduc.2022.942224

The focus and timing of gaze matters: Investigating collaborative knowledge construction in a simulation-based environment by combined video and eye tracking

Joni Lämsä[1,2]*, Jimi Kotkajuuri[3], Antti Lehtinen[4,5], Pekka Koskinen[4], Terhi Mäntylä[5], Jasmin Kilpeläinen[5] and Raija Hämäläinen[1]

[1]Department of Education, University of Jyväskylä, Jyväskylä, Finland, [2]Learning and Educational Technology Research Unit, University of Oulu, Oulu, Finland, [3]Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland, [4]Department of Physics, University of Jyväskylä, Jyväskylä, Finland, [5]Department of Teacher Education, University of Jyväskylä, Jyväskylä, Finland

Although eye tracking has been successfully used in science education research, exploiting its potential in collaborative knowledge construction has remained sporadic. This article presents a novel approach for studying collaborative knowledge construction in a simulation-based environment by combining both the spatial and temporal dimensions of eye-tracking data with video data. For this purpose, we have investigated two undergraduate physics student pairs solving an electrostatics problem in a simulation-based environment *via* Zoom. The analysis of the video data of the students' conversations focused on the different collaborative knowledge construction levels (new idea, explication, evaluation, and non-content-related talk and silent moments), along with the temporal visualizations of the collaborative knowledge construction processes. The eye-tracking data of the students' gaze, as analyzed by epistemic network analysis, focused on the pairs' spatial and temporal gaze behavior. We illustrate how gaze behavior can shed light on collaborative knowledge construction in terms of the quantity of the talk (e.g., gaze behavior can shed light on the different activities of the pairs during the silent moments), quality of the talk (e.g., gaze behavior can shed light on the different approaches when constructing knowledge on physical phenomena), and temporality of collaborative knowledge construction processes [e.g., gaze behavior can shed light on (the lack of) attempts to acquire the supporting or contrasting evidence on the initial ideas on the physical phenomena]. We also discuss the possibilities and limitations of gaze behavior to reveal the critical moments in the collaborative knowledge construction processes.

KEYWORDS

collaborative learning, epistemic network analysis, eye tracking, gaze, collaborative knowledge construction, multimodal data, simulation, video

Introduction

In science education, simulation-based environments have been used to foster collaborative knowledge construction (CKC) and guide students in building on each other's ideas and thoughts while learning about scientific phenomena (Schellens and Valcke, 2005; Liu et al., 2021). However, productive CKC processes in these environments rarely occur automatically (Jeong et al., 2019). Even though the (automatic) analysis of students' verbal conversations could provide information to teachers and machines so that they can guide CKC processes (Lämsä et al., 2021b), many nontrivial issues, such as moving from the retrospective modeling of learning processes to predictive analytics, must be solved before these applications can be more broadly adopted (Schneider et al., 2021). In the field of multimodal learning analytics, various data modalities are combined to comprehensively understand if and how learning occurs (Olsen et al., 2020). Ultimately, the aim is to use this information to support learning.

Collaborative knowledge construction analysis has typically focused on the quantity and quality of conversations *via* coding the utterances of video data and evaluation of learning outcomes (Jeong et al., 2014, 2019); in this context, the temporal analysis of CKC has gained increased attention (Lämsä et al., 2021a). In addition to the conversations captured with video data, CKC research could benefit from eye tracking (Olsen et al., 2020). Although eye tracking has been successfully used in science education research (e.g., Hahn and Klein, 2022), only a few studies have investigated the role of gaze in CKC in science learning (for an exception, see Becker et al., 2021). Gaze similarity among students has been associated with higher-quality learning processes and outcomes (Olsen et al., 2020; Becker et al., 2021), although exceptions do exist (Liu et al., 2021). On the one hand, it is essential to develop improved gaze similarity indicators that reveal both the focus and timing of gaze and, thus, better reflect the kinematics of CKC in simulation-based environments. On the other hand, the contextual information of CKC processes (such as video data) and eye-tracking data would ensure that the processes are interpreted reliably (Molenaar, 2021).

In the current article, we introduce a methodology of combined video and eye-tracking data analysis to study CKC kinematics in a simulation-based environment. We discuss the possibilities of this methodology for designing pedagogical practices for science education that can help us in understanding and guiding these CKC processes.

Literature review

Collaborative knowledge construction in simulation-based environments

Simulations work to model activities and processes by omitting irrelevant variables from the perspective of learning goals (Chernikova et al., 2020). Moreover, simulations provide users with certain control when they accomplish a given task (Chernikova et al., 2020). Simulations have been applied in practicing authentic practices and procedures, for example, in the aviation (Mavin et al., 2018) and healthcare sectors (Cook et al., 2013), and in various disciplines as a part of formal education (Chernikova et al., 2020). Simulations can be used in nondigital settings (e.g., simple patient simulations in healthcare) or digital settings (e.g., virtual reality flight simulations in aviation). In the current study, we focus on computer simulations in science education. In the context of science education, computer simulations are programs that provide a representation of a scientific phenomenon through a model (Clark et al., 2009; de Jong and Lazonder, 2014). Computer simulations such as PhET (University of Colorado Boulder, 2022b) or WISE (UC Berkeley, 2022) may improve learning outcomes by enhancing other forms of instruction, such as lectures or laboratories (Rutten et al., 2012; de Jong et al., 2013). Moreover, simulations may facilitate collaboration among students during CKC processes (Lämsä et al., 2018, 2020). This collaboration among students may be beneficial for gaining conceptual and procedural knowledge (Jensen and Lawson, 2011; Rutten et al., 2012).

In simulation-based environments, the user can interact with the simulation by exploring the effects of the given input variables to observe the effects on the output variables (Clark et al., 2009; de Jong and Lazonder, 2014). Within computer-mediated settings, the interaction between the students and simulation may take different forms (Figure 1). First, when using individual-based simulations, each student can individually interact with the simulation, requiring intensive verbal coordination of CKC processes between students (Figure 1A), which can be a challenge (Chang et al., 2017). Second, when using collaborative simulations, the students can interact with the simulation in a shared space collaboratively; hence, the coordination of CKC processes may be further fostered by assigning students distinct responsibilities (Figure 1B). Even though these latter simulation-based environments may benefit from the coordination of the CKC processes and, thus, facilitate interactions among students, they do not necessarily lead to higher-level CKC processes or learning outcomes compared with the former settings (Chang et al., 2017; Liu et al., 2021). Third, the

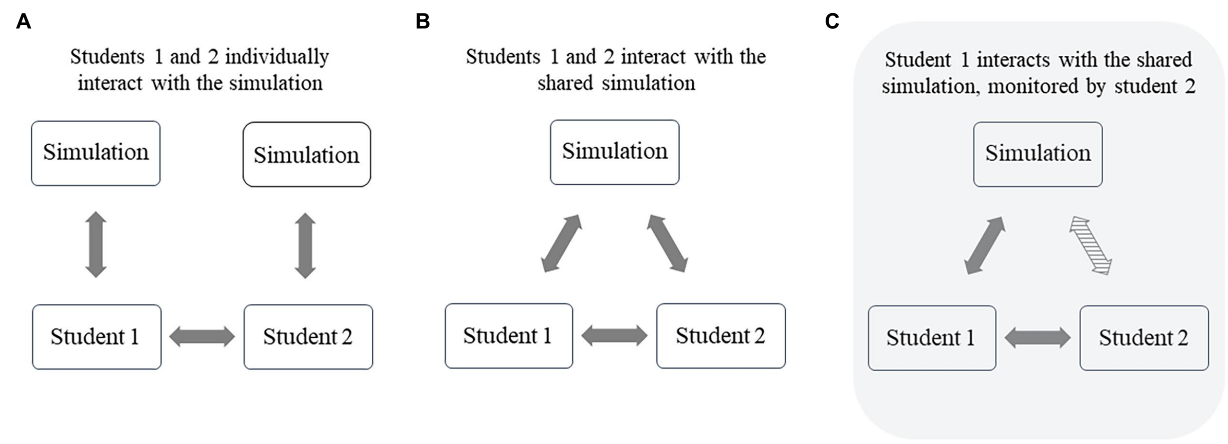


FIGURE 1
The interactions (double-sided arrows) among the students and simulation with the **(A)** individual-based simulations, **(B)** collaborative simulations, and **(C)** individual-based simulations used with the screen-sharing functionality in computer-mediated collaborative knowledge construction. The current study focuses on the faded scenario **(C)**, in which student 1 is sharing and student 2 is monitoring the screen.

rapid adoption of communication apps, especially during the COVID-19 lockdown, shifted face-to-face sessions to Zoom or Teams. Sessions with screen sharing can also foster the coordination of the CKC process, even if individual-based simulations are used (Stevenson et al., 2022). Although one student sharing the screen interacts with the simulation and others monitor the simulation view from the screen, the sharing student can mediate the interaction between the simulation and others by implementing requests from others in the simulation environment (Figure 1C). In the current study, we focus on this third scenario.

Although this scenario inevitably assigns different roles to the students in the CKC processes, both sharing and monitoring students should effectively utilize a simulation-based environment as an external resource for explicating and evaluating ideas and thoughts (Jeong and Hmelo-Silver, 2010). Usually, students share their ideas and thoughts without building on previous ones, meaning that critical explication and evaluation of others' ideas and thoughts and other higher levels of CKC are rare (Yang et al., 2018). Students may also have challenges understanding visual representations of abstract concepts, such as fields (Klein et al., 2018). These challenges highlight the role of the teacher and simulation-based environment in guiding CKC (Lin et al., 2013; Lehtinen and Viiri, 2017). In this respect, the eye-tracking analysis provides a view of students' visual attention that could provide teachers and simulation developers with information on unnecessary or distracting visual objects, helping guide CKC and improve these environments.

Studying collaborative knowledge construction with eye tracking

In the current paper, we refer to gaze as "the act of directing the eyes toward a location in the visual world" (Hessels, 2020, p. 856) and gaze behavior as gaze similarities and dissimilarities over time. Tatler et al. (2014, p. 6) have pointed out that "eye movements give us a window onto how perception operates across the course of a task, from the first intention to act and through the process of carrying out the task itself." Strohmaier et al. (2020) showed that many studies using eye tracking to study learning processes assume that when a student's gaze is focused on an artifact, the student processes the information being provided (see Just and Carpenter, 1980). This assumption, however, is a simplification because, even though the sharp image of the artifact is formed within a tiny area of the eye, which is called the fovea (Holmqvist et al., 2011, p. 21), humans can process information from the wider area around the artifact (parafoveal processing, Schotter et al., 2012).

One of the critical questions in CKC is how to capture a joint activity between pairs or small groups using eye tracking (e.g., Hayashi and Shimojo, 2021). So far, most studies have evaluated CKC processes by assessing how often students look at the same objects of the learning environment (Olsen et al., 2020; Becker et al., 2021; Sharma et al., 2021). For example, Becker et al. (2021) found that early gaze similarities concerning laboratory apparatus were positively associated with the learning outcomes in a collaborative laboratory. However, similar gaze patterns do not guarantee productive learning processes and outcomes (Schneider et al., 2018). For example, high gaze similarity may result in low-level CKC processes and poor outcomes if the similarity is related to irrelevant objects (a synthesis by Hahn and Klein, 2022, indicated this to be true when learning individually in simulation-based environments). Schneider et al. (2018) addressed this challenge in the literature by augmenting spatial information from eye-tracking data and verbal information from audio recordings into cross-recurrence graphs that indicate "how and the extent to which streams of information come to exhibit similar patterns in time" (Coco and Dale, 2014, p. 2). Simulations often visualize concepts that are abstract, nonlocal, and visually absent in the real world, such as fields and forces. Thus, rich visualizations result in several visual objects, which can complicate the interpretation and comparison of cross-recurrence graphs.

The analysis of students' gaze behavior means identifying the temporal co-occurrences of their gaze events (see an overview of the temporal analysis methods in Lämsä et al., 2021a). For this purpose, an emerging method in the learning sciences is epistemic network analysis (ENA; Shaffer et al., 2016). The premise of ENA is that co-occurrences of (gaze) events are more important than the events as such (Shaffer et al., 2016; Andrist et al., 2018). ENA models the co-occurrences of the gaze events with nodes and edges: the areas of interest (AOIs) are depicted as nodes, and the co-occurrences of the students' gaze events with these AOIs are depicted as the edges between nodes. An advantage of the ENA compared with other network analysis methods is that it allows for examining which (instead of how) nodes are connected (Bowman et al., 2021); from the perspective of CKC processes, this is important to understand which features of the simulation-based environment students are simultaneously looking at. Moreover, the ENA allows for comparisons of the networks by keeping the nodes and edges in the same location in the visualization of the networks (Bowman et al., 2021); this facilitates a comparison of the students' gaze behaviors between the pairs or small groups and between the CKC levels.

In the current study, we introduce a novel approach for exploring CKC kinematics in a simulation-based environment. By kinematics, we refer to the connections between the CKC processes and gaze behavior without considering their dynamics, which would imply understanding the causes of the observed CKC processes or gaze behavior. To illustrate our approach, we use video data of student pairs' conversations to understand their CKC processes from the perspectives of the (i) quantity of the talk, (ii) quality of the talk, and (iii) temporality. We then apply ENA to eye-tracking data to explore what insights the student pairs' gaze behavior provides regarding these CKC processes. We answer the following research questions (RQs):

RQ1: What does the analysis of the video data tell about the pairs' CKC processes?

RQ2: What does the pairs' gaze behavior tell about these CKC processes?

Materials and methods

Context and participants

The current study was conducted in an introductory electricity course at a Finnish university. We focus on the data from two student pairs who used the *Charges and Fields* PhET simulation (University of Colorado Boulder, 2022a) to solve an electricity problem (Figure 2). The students worked in different rooms *via* Zoom so that both saw the same assignment and simulation views of the split screen. One student shared (S) and the other monitored (M) the screen; in the rest of this paper, we refer to these students as 1S and 1M (the sharing student and monitoring student of pair 1, respectively), and 2S and 2M (the sharing student and monitoring student of pair 2, respectively). The pairs constructed knowledge of electric field properties in the presence of a static negative charge and positive charge that could be moved about. The students were supposed to apply the superposition principle to explain how the direction and magnitude of the nonlocal electric field change when moving the positive charge.

Data

To answer RQ1, we video recorded the pairs' conversations in Zoom and transcribed the pairs' conversations (pair 1: 5.3 min. with 59 utterances; pair 2: 12.8 min. with 163 utterances) using the "unit of meaning" (Henri, 1992, p. 134) to identify episodes comprising a few utterances. We then applied Veerman and Veldhuis-Diermanse (2001) to analyze the CKC level through theory-driven content analysis. We coded the episodes (13 and 38 episodes) as either *physics content–related talk*, including the following CKC levels: (i) new idea, (ii) explication (elaboration on earlier ideas), and (iii) evaluation (critical discussion of and reasoning about earlier ideas), or *non-content-related talk*, including planning and technical talk (e.g., planning procedures or wondering how to invoke the simulation). The first author prepared a coding manual with the definitions and example excerpts of the codes. After this, the first author and the coauthor coded all the episodes of the pairs' conversations, after which the disagreements (see the contingency table in Table 1) were resolved and definitions of the codes revised by all the authors (see Table 2).

To answer RQ2, we collected the eye-tracking data using Tobii Pro Glasses 2 (sampling frequency 50 Hz), which are mobile wearable eye trackers. Eye tracking allowed free movement of the participants so that the gaze outside the computer screen could also be captured. The scene camera of the eye tracker had a resolution of 1,920 × 1,080 pixels, capturing 52° vertically and 82° horizontally. We used one-point calibration, and we verified the calibration by asking the participants to look at three different points on their surroundings (the points were left, right, and in front of them). We wanted to keep the data collection situation as authentic a learning situation as possible, and we did not use chinrests, control students' distance to the computer screen, nor control the gaze angles; however, the learning situation and simulation-based environment (Figure 2) provided satisfactory conditions for eye-tracking data collection (e.g., distance to the computer screen was approximately 0.5–1.0 m, and the targets in the environment were located within narrow area so that no large gaze angles were needed that improved the accuracy of the eye tracking; Tobii Pro AB, 2017). The data were analyzed in Tobii Pro Lab (Tobii Pro AB, 2022). We used the Tobii I-VT (Attention) as a gaze filter, which is the default preset for wearable eye trackers. The velocity threshold parameter was 100°/second. Blinks and saccades were cleaned from the data, and only fixation data were used in the coding and further analyses.

10.3389/feduc.2022.942224

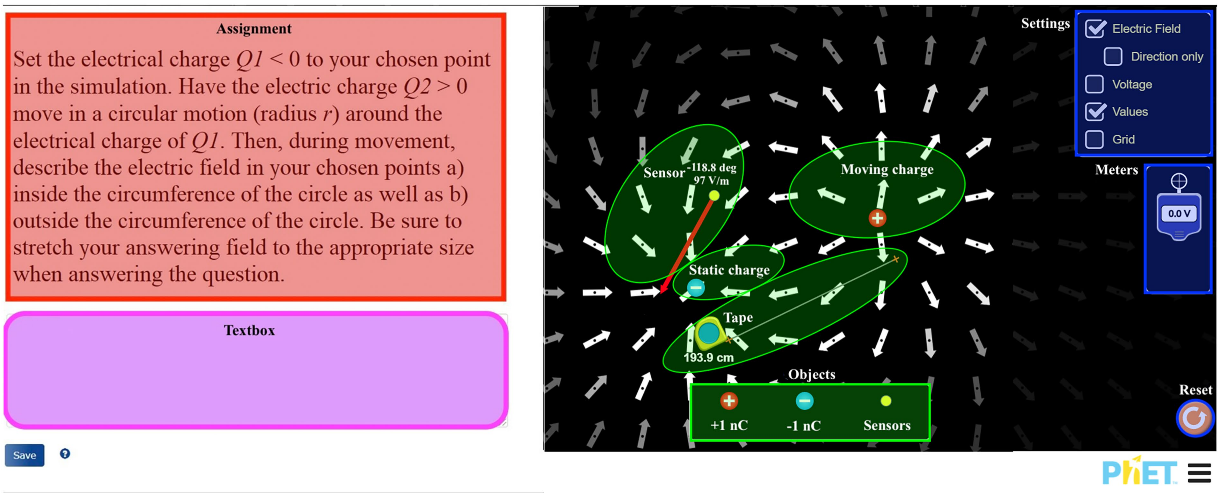


FIGURE 2
The assignment and simulation view the student pairs were looking at when they constructed knowledge on electric field properties in the presence of a negative static charge and positive movable charge. The areas of interest are labeled using colored shapes; the labels were not visible to the students. The students wrote their answers to the problem in the textbox on the left.

TABLE 1 The contingency table shows the agreements and disagreements in the coding of the conversations between two coders.

		Coder 1			
		Non-content-related	New idea	Explication	Evaluation
Coder 2	Non-content-related	27	0	0	0
	New idea	0	7	0	0
	Explication	1	0	12	1
	Evaluation	0	0	0	3

All 51 episodes were coded by two authors.

To study the gaze behavior, we first watched eye-tracking recordings to explore where students divided their visual attention when solving the given problem. Based on this exploration and the expert analysis of the problem itself (five authors have master's or doctoral degrees from physics), we divided the screen view into AOIs (see Figure 2 and Table 3; the keyboard was an AOI only for the sharing student). The formed AOIs allowed "local analysis" (Hahn and Klein, 2022, p. 5), which differentiates the irrelevant and relevant features of the simulation view (Figure 2 and Table 3). The fixation data were manually coded in Tobii Pro Lab into the different AOIs based on the screen capture in Figure 2. The coding was done fixation by fixation, clicking that AOI in the screen capture to which the student's gaze was located in the eye-tracking recording. The coding decisions were made based on the set of objects in the screen capture (e.g., sensor, moving charge, and static charge), not on the absolute position of the fixation in the eye-tracking recording (e.g., if the student's visual attention was on the moving charge, it was coded as such, even though the position of the moving charge in the eye-tracking recording would have differed from that presented in Figure 2). Fixations unrelated to any AOIs were coded as being "outside screen" and excluded from further analysis. Two researchers coded the fixation data of one student (562 fixations, of which 160 were "outside screen"). To check the interrater reliability of the coding, we then calculated Cohen's kappa (Cohen, 1960) and Shaffer's rho separately for each code (AOI; Table 3) so that a high agreement in one code did not hide a low agreement in another code (Shaffer, 2017; Eagan et al., 2020). Cohen's kappa was >0.97 for all the codes (AOIs), indicating almost perfect agreement between the two coders (Shaffer's rho was <0.05 for all the codes when we set 0.7 as a threshold value of Cohen's kappa to indicate good reliability; Cicchetti, 1994).

Because the sampling frequency of the eye trackers was 50 Hz (a data point for each 20-ms interval), as a result, we had a time series of the gaze events in which all AOIs were assigned binary data for each 20-ms interval, here corresponding either to student visual attention (one) or the absence of student visual attention (zero). We excluded five AOIs (settings, objects, measuring tape, meters, and reset) because they rarely attracted students' attention (Table 3); this exclusion also eased the interpretation of the epistemic networks by decreasing the number of nodes in the networks. For both student pairs, synchronization of the video and eye-tracking data enabled analysis of the CKC processes from the perspectives of the (i) quantity of the talk, (ii) quality of the talk, and (iii) temporality and gaze behavior.

Analysis

To answer RQ1 and to study the quantity of the talk, we first calculated the relative amount of time that the pairs used for non-content-related talk and physics content-related talk, including the following CKC levels: (i) new idea, (ii) explication, and (iii) evaluation. We also calculated the relative amount of time

TABLE 2 Coding manual for non-content-related talk and physics content–related talk that includes a code for each level of collaborative knowledge construction.

Content	Code	Definition	Example excerpt
Non-content-related talk	Planning, coordination and technical talk	Planning and coordinating procedures or wondering how to invoke the simulation	2M: Now we just write down very neatly that the electric fie … Hang on a second, the electric field itself … 2S: Yeah, so what was the question? Descr … the electric field …
Physics content–related talk	CKC: New idea	Presenting a new idea or thought in the context of the ongoing conversation	2S: Yes, it [the electric field] is at the smallest when it is here on the oppos … other side.
Physics content–related talk	CKC: Explication	Elaborating further on earlier ideas	2M: Yes, then at the largest when they are so that one [charge] is attracting it and another [charge] is pulling it in the same direction. Yes.
Physics content–related talk	CKC: Evaluation	Discussing critically and reasoning about earlier ideas	2M: Mm, while approaching how does it … [commenting on the written answer in the textbox] 2S: Or it can … What? 2M: How is it approaching … That is, while approaching? 2S: So, it is here like that. Here, when it's farther away, and then, when it is approaching there, then that force starts to increase. 2M: Mm, okay. But does it increase when it is on the side, even though it is already approaching?

2S and 2M refer to the sharing student and monitoring student of pair 2.

TABLE 3 The areas of interest (AOIs) and their total fixation durations in percentages during the collaborative knowledge construction processes of pair 1 (1S and 1M, duration 5.3 min) and pair 2 (2S and 2M, duration 12.8 min).

AOI/Student	1S (%)	1M (%)	2S (%)	2M (%)
Assignment	13.0	13.6	7.4	3.5
Textbox	16.8	32.1	10.9	4.4
Keyboard	20.5	–	2.6	–
Sensor	4.2	0.4	17.8	15.6
Moving charge	5.8	1.5	12.7	6.8
Static charge	5.6	13.8	14.3	2.9
Settings	1.0	0.8	0.5	0.1
Objects	0.6	0.5	0.8	0.3
Measuring tape	3.3	4.9	0.0	0.0
Meters	0.6	0.1	0.4	0.0
Reset	0.0	0.0	0.0	0.0
Total	**72.4**	**74.1**	**70.1**	**46.5**

S and M refer to the sharing student and monitoring student, respectively. The shaded AOIs are used in further analyses, and the dashed line separates the AOIs referring to the problem and simulation view.

for silent moments. Second, to study the quality of the talk, we examined the quality of the conversations at the different CKC levels in terms of whether students' ideas (and explication and evaluation of those ideas) were correct or not in the context of the given problem (Figure 2). Third, we studied the temporality of the pairs' CKC processes by visualizing CKC level and non-content-related talk as a function of time.

To answer RQ2, we applied ENA to synchronized, binary eye-tracking data (see Shaffer et al., 2016; Andrist et al., 2018). The AOIs served as the nodes of the network (Figure 2). We considered that the gaze events within a 2-s time interval were connected, so we used the moving windows of the size of a 100 rows (100 rows × 20 ms/row = 2 s). We chose this 2-s time interval based on previous studies on the gaze similarity of pairs (Richardson and Dale, 2005; Schneider et al., 2018). The unit of analysis was a pair at different CKC levels (along with during non-content-related talk and during silent moments), so we created the adjacency matrices for both pairs at each CKC level separately. The adjacency matrices represent the strength of the connections between the AOIs of the two students at the different CKC levels (along with during non-content-related conversations and during silent moments). We used weighted sums so that more connections between the AOIs within a moving window also resulted in stronger connections between these AOIs. When building epistemic networks, we did not visualize the connections between the AOIs of an individual student; in other words, if the student focused on several AOIs within the 2-s time interval, the connections between these AOIs were not visible in the epistemic networks (Andrist et al., 2018). We made this decision to facilitate the interpretation of the networks.

After the adjacency matrices for each unit of analysis had been created, the matrices were converted into adjacency vectors (Bowman et al., 2021) that were spherically normalized. This normalization eased the comparison of the networks when the duration of the CKC processes (and, thus, the number of gaze events) differed between pairs 1 and 2 (see Table 3). Finally, the dimensions of the adjacency vectors were reduced by singular value decomposition, after which the network nodes were positioned by applying an optimization method (see Bowman et al., 2021). The networks included two nodes for each AOI (see Table 3): one node for the sharing student and another for the monitoring student. The edges connecting the nodes provided a visualization of the gaze behavior: the thicker the edge, the more students had simultaneously focused on the corresponding AOIs within the two-second time interval. Figure 3 demonstrates this process with a fictional,

A

Time interval (moving window) 1

	AOI1_S	AOI1_M	AOI2_S	AOI2_M
AOI1_S		4	2	0
AOI1_M	4		4	1
AOI2_S	2	4		0
AOI2_M	0	1	0	

Time interval (moving window) 2

	AOI1_S	AOI1_M	AOI2_S	AOI2_M
AOI1_S		2	3	0
AOI1_M	2		8	5
AOI2_S	3	8		0
AOI2_M	0	5	0	

B

	AOI1_S	AOI1_M	AOI2_S	AOI2_M
AOI1_S		6	5	0
AOI1_M	6		12	6
AOI2_S	5	12		0
AOI2_M	0	6	0	

C

FIGURE 3
A fictional, simplified process for visualizing the epistemic network of two areas of interest (AOI1 and AOI2). *S* and *M* after the underscore refer to the gaze of the sharing student and monitoring student. **(A)** Weighted adjacency matrices represent the co-occurrences of gaze events in the two different time intervals (moving windows). **(B)** The cumulative adjacency matrix is calculated by summing the adjacency matrices presented in **(A)**. The connections in the AOIs between the sharing and monitoring students are included in further analyses; the connections within an individual student are excluded (see the shadings in the matrices). **(C)** The arbitrary visualization of the epistemic network shows that the connections between AOI2_S and AOI1_M were stronger (a thicker edge between the nodes) than the connections between AOI1_S and AOI1_M (a thinner edge between the nodes). The sharing student focused their attention neither on the AOI1_S nor on the AOI2_S when the monitoring student was looking at AOI2_M.

simplified dataset. We performed the ENA in RStudio (Version 1.2.1335) by applying the rENA package (Marquart et al., 2021).

Results

In the following section, we cover the pairs' CKC processes from the perspectives of the (i) quantity of the talk, (ii) quality of the talk, and (iii) temporality based on the analysis of the video data (RQ1, Section "Pairs' collaborative knowledge construction processes based on the video data"). We then illustrate what insights the pairs' gaze behavior provides regarding these CKC processes (RQ2, Section "Pairs' collaborative knowledge construction processes: Insights based on gaze behavior").

Pairs' collaborative knowledge construction processes based on the video data

Quantity of the talk

Figure 4 shows the relative amount of time that the pairs used for non-content-related talk and physics content–related talk, including the following CKC levels: (i) new idea, (ii) explication, and (iii) evaluation. The relative amount of time for silent moments is also shown in Figure 4. Pair 1 had more silent moments and less non-content-related talk, such as planning, than pair 2 (67% vs. 42 and 15% vs. 25%, respectively). Regarding the physics content–related talk, both pairs used a relatively similar amount of time to present new ideas (6 and 6%) and explicated those (13 and 17%), but pair 2 also evaluated the presented ideas 10% of the time.

Quality of physics content–related talk

Even though there were no differences between the pairs in the relative amount of physics content–related talk in presenting new ideas and explicating those (Figure 4), pair 1 exhibited low quality of physics content–related talk. The new ideas that the monitoring student (1M) presented to the problem did not include the magnitude of the electric field, instead focusing only on its direction. These ideas about the direction were also incorrect because 1M ignored the fact that the direction of the electric field was constantly changing when the positive charge was moved (starting time of the utterance at $t = 1.9$ min, see Figure 5A):

> 1M (Monitoring student): Well, inside those [the electric field lines], all of them are pointing toward the negative [static charge].

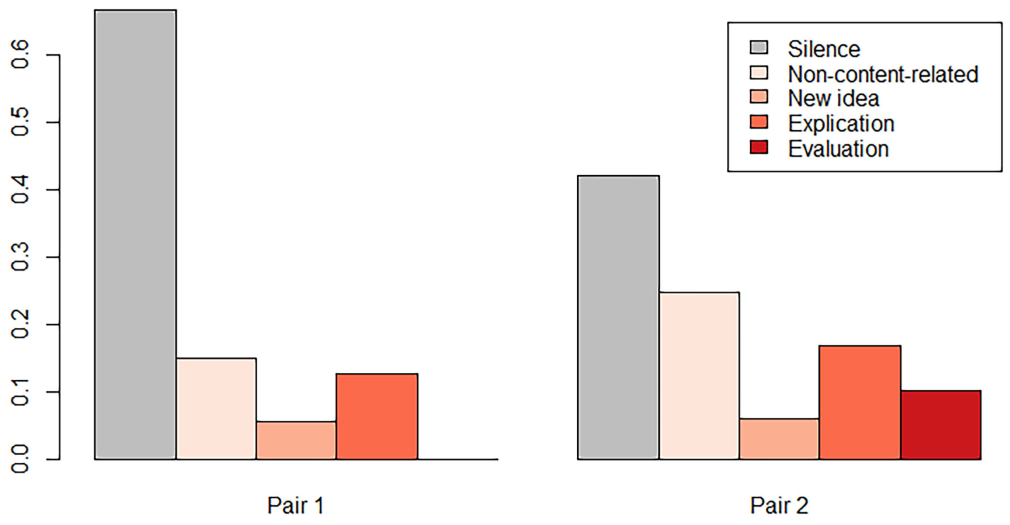


FIGURE 4
The relative amount of talk (in %) at the different collaborative knowledge construction (CKC) levels. The amount of non-content-related talk and silent moments has also been marked.

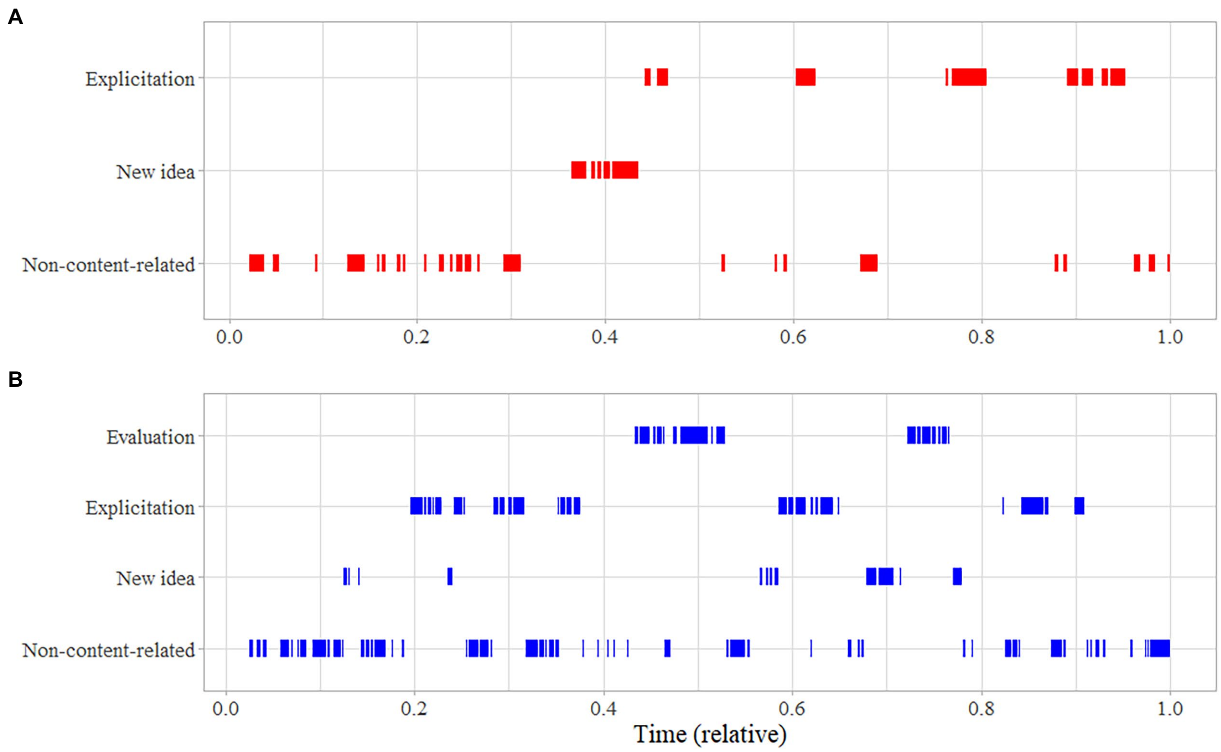


FIGURE 5
Visualization of the collaborative knowledge construction (CKC) process of **(A)** pair 1 and **(B)** pair 2. The duration of the CKC process was 5.3 min for pair 1 and 12.8 min for pair 2.

1S (Sharing student): Mm. Yes … And outside then … But does [the electric field] change if … Mm.

1M: Yes, so then it’s kind of … There, where the positive [moving charge] is, so then those [the electric field lines] are pointing away from its vicinity, but otherwise, it is always pointing toward the negative [static charge].

1S: Yes.

Later, pair 1 only pondered and explicated how 1M's incorrect ideas could be formulated to write in the textbox (starting time at $t = 2.3$ min; see Figure 5A):

> 1S: So, hmm.
>
> 1M: For a), all [the electric field lines] point toward the negative [static charge].
>
> 1S: Yes.

In contrast, the monitoring student of pair 2 (2M) presented fair ideas of the problem, even though 2M also focused more on the direction of the electric field than its magnitude (starting time at $t = 1.4$ min, see Figure 5B):

> 2M (Monitoring student): But it [the electric field] is doing that kind of pendular motion there.
>
> 2S (Sharing student): So it is. Yeah.

In the explication level, 2M provided physical explanations for the presented ideas and thoughts (starting time at $t = 2.3$ min; see Figure 5B):

> 2M: Let's write this neatly down so that the direction of the force starts to oscillate then and … Then in part a), inside … Hmm … The direction of the [electric] field is changing, of course, depending on their lengths. Or no, depending on the … Hmm … Kind of position where the moving charge is going. Thus, a kind of oscillatory motion emerges. Because it is rotating 180° or, ahem, pi radians, it is always on the side where they kind of constructively interfere and half of which are destructive.

Pair 2 also evaluated the presented ideas (see an example in Table 2), while this CKC level was absent in pair 1's conversation.

Temporality of CKC processes

Figure 5A shows that the CKC process of pair 1 moved straightforwardly from non-content-related talk to presenting new ideas and then to explication without evaluation. Non-content-related talk, including planning and coordinating actions as examples, was rare later in the CKC process (see Section "Quantity of the talk"). Thus, pair 1 made their conclusions based on their initial and incorrect ideas and thoughts (see Section "Quality of physics content–related talk"), which they only explicated on further (no transitions from explication to presenting new ideas). Thus, pair 1 failed to solve the problem shown in Figure 2 correctly because they concluded the following in their joint answer in the textbox:

> The electric field inside the circumference of the circle always points toward the [static] charge Q_1.
>
> [Outside the circumference of the circle and] close to the [moving] charge Q_2, the electric field points away. When charge Q_2 moves further, the electric field again points toward the [static] negative charge.

The answer reveals that pair 1 did pay attention to the direction of the electric field but not to its magnitude. They also failed to notice how the direction of the electric field constantly changed when the positive charge moved around the negative charge.

Figure 5B shows that pair 2 had several transitions between the CKC levels and non-content-related talk, meaning that pair 2 frequently planned their actions (Section "Quantity of the talk"). These findings may relate to their problem-solving strategy, which separately considered the two aspects of the problem: the electric field inside (0–7 min) and outside the circle (7–13 min, Figure 5B; see also Figure 2). Pair 2 reached the highest CKC level when they evaluated their ideas in both parts of the problem. Pair 2 finally focused both on the magnitude and direction of the electric field, answering the problem more correctly:

> [Inside the circumference of the circle], the direction [of the electric field] changes periodically; [and] the magnitude [of the electric field] increases when the [moving] charge Q_2 approaches the [chosen] point a.
>
> [Outside the circumference of the circle], when the [moving] charge Q_2 is on the same side of the circumference of the circle as the [chosen] point a, the [electric] fields add up.

The answer illustrates that pair 2 made relevant observations on electric field properties, despite a few careless statements, such as that electric fields add up only under certain conditions ("when the [moving] charge Q2 is on the same side of the circumference of the circle as the [chosen] point a"). We now explore what kinds of insights the pairs' gaze behavior provides on these three perspectives of the CKC processes that we covered in sections "Quantity of the talk", "Quality of physics content–related talk", and "Temporality of CKC processes".

Pairs' collaborative knowledge construction processes: Insights based on gaze behavior

Gaze behavior sheds light on the silent moments and non-content-related talk

First, the pairs' gaze behavior reveals that the silent moments had different purposes from the perspective of CKC (see section "Quantity of the talk"): Figure 6A indicates that pair 1 used these silent moments for writing their answer to the textbox (1S's visual attention was on the keyboard, while 1 M's visual attention was on the textbox). Figure 6B shows that pair 2 used these silent

10.3389/feduc.2022.942224

moments for working with the simulation (1S's visual attention was on the sensor, moving charge, and static charge, while 1 M's visual attention was on the sensor), and both students also focused their visual attention on the textbox. The difference between these two networks is presented in Figure 6C, indicating that pair 1 used more time for formulating their answer to the textbox and less for working with the simulation than pair 2.

Second, the pairs' gaze behavior indicates that the CKC processes during the non-content-related talk differed between the pairs, as was the case with the silent moments. The pairs' gaze behavior in Figure 7 shows that the students in pair 1 paid more visual attention to the assignment and textbox than the students in pair 2 (Figures 7A,C). The students in pair 2 divided their visual attention more on the simulation view than the students in pair 1 (Figures 7B,C).

Gaze behavior sheds light on the knowledge construction approaches

In section "Quality of physics content–related talk," we found that pair 1 did not present correct ideas about the direction of the electric field, and both pairs ignored the magnitude of the electric field at the beginning of their CKC processes. When presenting

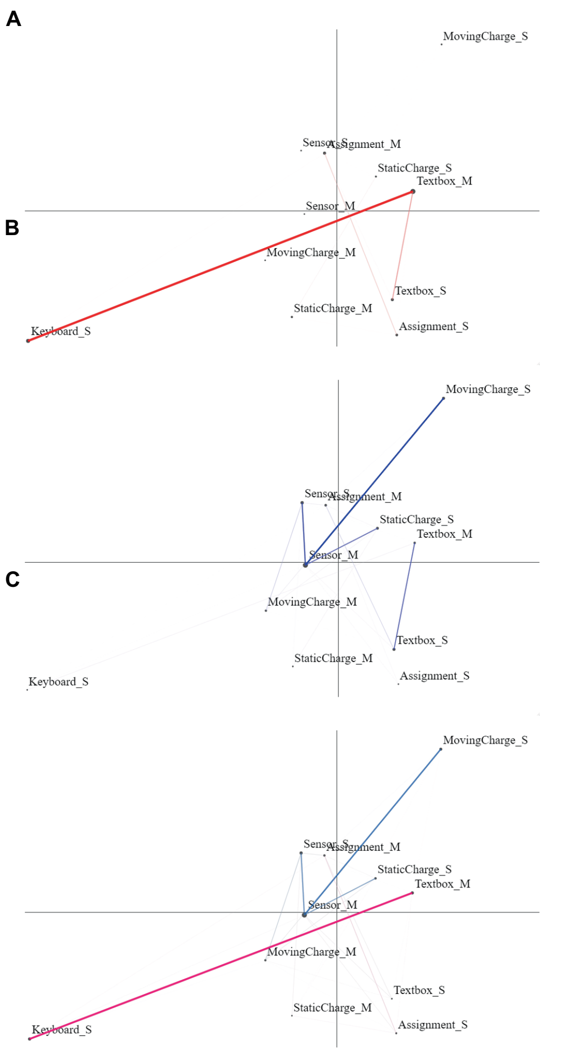


FIGURE 6
Epistemic networks of the gaze behavior of **(A)** pair 1 and **(B)** pair 2 during silent moments. *S* and *M* after the underscore refer to the gaze of the sharing student (1S/2S) and the monitoring student (1M/2M). The difference between epistemic networks **(A,B)** is presented in **(C)**. The red edges show the connections between the nodes that were stronger among pair 1 than among pair 2. The blue edges show the connections between the nodes that were weaker among pair 1 than among pair 2.

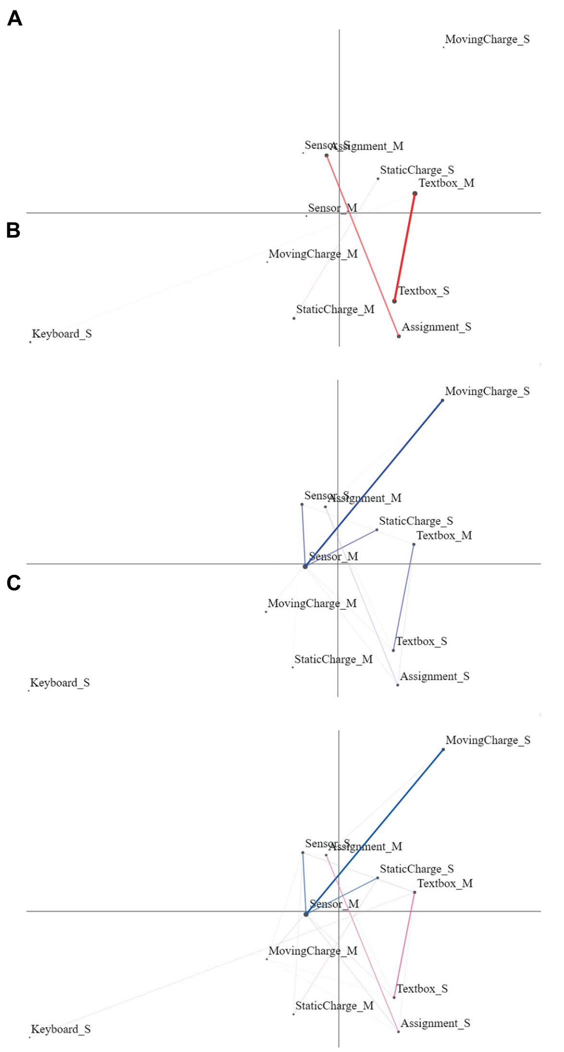


FIGURE 7
Epistemic networks of the gaze behavior of **(A)** pair 1 and **(B)** pair 2 during non-content-related talk. *S* and *M* after the underscore refer to the gaze of the sharing student (1S/2S) and monitoring student (1M/2M). The difference between epistemic networks **(A,B)** is presented in **(C)**. The red edges show the connections between the nodes that were stronger among pair 1 than pair 2. The blue edges show the connections between the nodes that were weaker among pair 1 than pair 2.

new ideas, pair 1 had gaze dissimilarities, so that both students paid attention to the moving charge but not simultaneously (Figure 8A). Pair 2 had gaze similarities, and they were both simultaneously paying visual attention to the moving charge (Figure 8B); these differences are also visible in the difference network in Figure 8C. It is remarkable that neither of the monitoring students paid attention to the sensor when they presented new ideas to the problem, even though the sensor provided information on the direction and magnitude of the electric field.

The pairs' gaze behavior during the explication shows their different approaches when constructing knowledge on the properties of the electric field. Figure 9A shows that both 1M and 1S focused on the textbox, with only a few fixations on the simulation view (note what we found in section "Quality of physics content–related talk": pair 1 explicated how 1M's incorrect ideas could be formulated in the textbox). In contrast, Figure 9B shows that 2S and 2M focused their attention on the sensor, while 2S also focused on the moving charge (note that pair 2 aimed to provide physical explanations of the presented ideas during the explication, as we found in section "Quality of physics content–related talk"). These differences between the pairs' gaze behaviors are also visible in the difference network in Figure 9C.

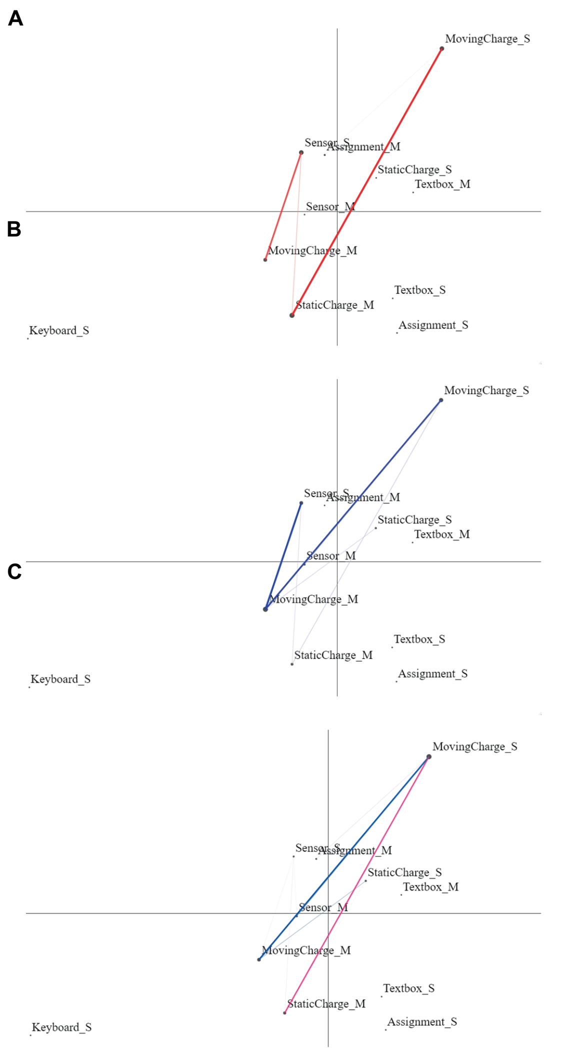


FIGURE 8
Epistemic networks of the gaze behavior of **(A)** pair 1 and **(B)** pair 2 when presenting new ideas. *S* and *M* after the underscore refer to the gaze of the sharing student (1S/2S) and monitoring student (1M/2M). The difference between epistemic networks **(A,B)** is presented in **(C)**. The red edges show the connections between the nodes that were stronger among pair 1 than among pair 2. The blue edges show the connections between the nodes that were weaker among pair 1 than among pair 2.

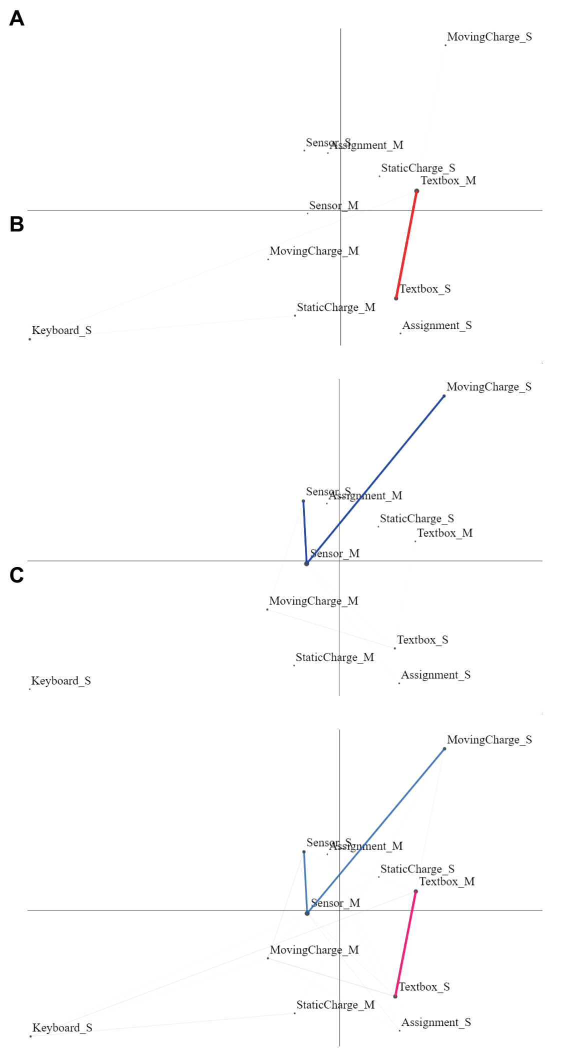


FIGURE 9
Epistemic networks of the gaze behavior of **(A)** pair 1 and **(B)** pair 2 during explication. *S* and *M* after the underscore refer to the gaze of the sharing student (1S/2S) and the monitoring student (1M/2M). The difference between epistemic networks **(A,B)** is presented in **(C)**. The red edges show the connections between the nodes that were stronger among pair 1 than among pair 2. The blue edges show the connections between the nodes that were weaker among pair 1 than among pair 2.

10.3389/feduc.2022.942224

Gaze behavior sheds light on the temporality of CKC processes

As we have seen in sections "Gaze behavior sheds light on the silent moments and non-content-related talk" and "Gaze behavior sheds light on the knowledge construction approaches", both students of pair 1 focused on the assignment, textbox, and keyboard, except during the short phase when they presented new ideas regarding the problem and focused on the simulation view (Figure 8A). This kind of gaze behavior implies that pair 1 had a few moments when they could have questioned the presented incorrect ideas to the problem; for example, monitoring student 1M hardly focused their visual attention on the sensor that provided information on the direction and magnitude of the electric field. Even though the sharing student (1S) focused their attention on the sensor when they presented new ideas, 1S did not question 1M's incorrect ideas about the problem (see section "Quality of physics content–related talk"). Based on the gaze behavior in the explication level (Figure 9A), neither 1M nor 1S tried to find supporting or contrasting evidence to the presented ideas because neither student consulted the simulation view during this CKC level.

Regarding pair 2, the edges (the blue lines) between the nodes (the AOIs) in Figures 6–10 show that pair 2 was more focused on the simulation view than pair 1 (particularly, see Figures 6C–10C). During the physics content–related talk, the visual attention of 1M and 1S was almost entirely on the simulation view (see new idea in Figure 8B, explication in Figure 9B, and evaluation in Figure 10). Pair 2 also used silent moments and non-content-related talk both for working with the simulation and formulating their solution to the problem in the textbox. This kind of gaze behavior constantly gave food for thought to the students (making new observations, explicating and evaluating those, and writing them down) that might be associated with the frequent transitions between the CKC levels and non-content-related talk that we found in section "Temporality of CKC processes."

Discussion

By combining video and eye-tracking data, we have introduced a novel approach to exploring CKC kinematics in a simulation-based environment. To illustrate our approach, we used video data of two student pairs' conversations to understand their CKC processes from the perspectives of the (i) quantity of the talk, (ii) quality of the talk, and (iii) temporality (RQ1). We then applied ENA to eye-tracking data to explore how gaze behavior can shed light on CKC processes in terms of these three perspectives (RQ2). As examples, we found that gaze behavior can shed light on (i) the learning activities of the pairs during the silent moments and non-content-related talk; (ii) the chosen approaches when constructing knowledge on physical phenomena; and (iii) (the lack of) attempts to acquire the supporting or contrasting evidence on the initial ideas on the physical phenomena.

Many studies have indicated that students' gaze similarities play a role in the learning processes and outcomes in collaborative learning settings (Schneider, 2019; Olsen et al., 2020; Becker et al., 2021). Our findings emphasize that instead of treating gaze

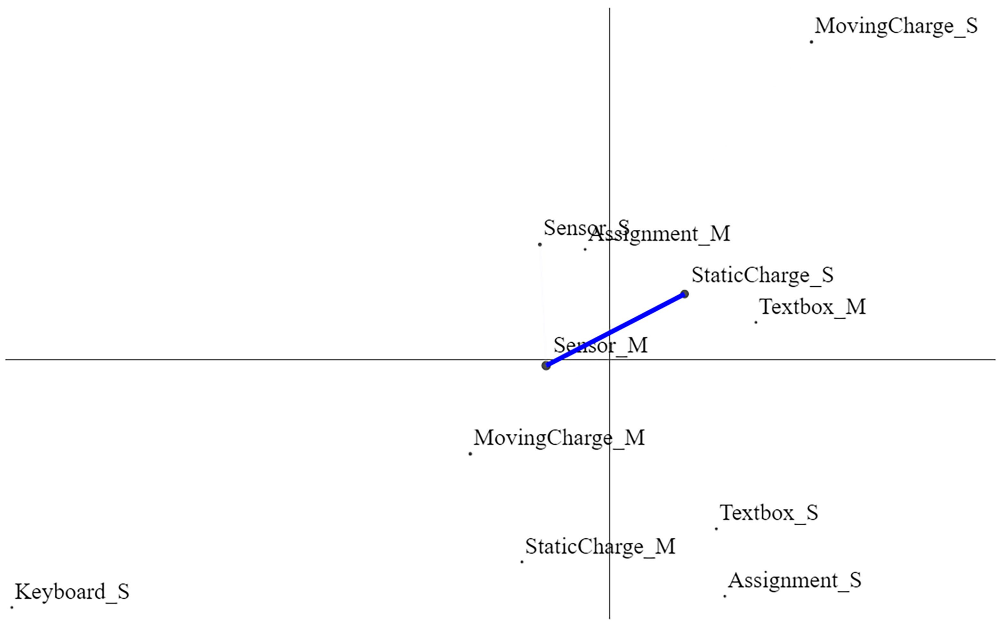


FIGURE 10
Epistemic networks of the gaze behavior of pair 2 during the evaluation (pair 1 did not evaluate their collaborative knowledge construction process). *S* and *M* after the underscore refer to the gaze of the sharing student (2S) and monitoring student (2M).

similarity merely as a binary variable and investigating the extent to which students are or are not looking at the same objects, comprehensive attention should be paid to investigating how gaze behavior can facilitate or hinder the ongoing CKC processes. In our study, pair 1 had a straightforward transition from presenting new ideas to explicating them (RQ1), and they hardly consulted the simulation view at the higher levels of their CKC process (RQ2, Figure 9). From the perspective of guiding students in their CKC processes, it is crucial to capture the critical moments of their CKC processes, such as the phase in which pair 1 presented new idea about the problem (Sections "Quality of physics content–related talk" and "Gaze behavior sheds light on the knowledge construction approaches", Figure 8). The analysis showed that student 1 M did not focus their visual attention on the sensor but only on the static and moving charges that did not provide information about the electric field. After that, pair 1 started the explication level by writing down their ideas and thoughts in the textbox, here without critical explication and evaluation of the presented ideas (Sections "Temporality of CKC processes" and "Gaze behavior sheds light on the temporality of CKC processes", Figure 9). As a form of guidance in these situations, students could be made aware of each other's gaze behavior and prompted to focus their visual attention on the relevant features of the simulation (Hayashi, 2020). The information on students' gaze behavior and its relation to the CKC process can also help teachers and developers of educational technology, for example, in how to visualize abstract concepts, such as fields, so that the selected representations can be effectively utilized as external resources of learning (Klein et al., 2018).

When considering the gaze behavior of the pairs, attention should also be paid to student roles during the CKC process. In our study, both pairs had one student sharing the screen and another student monitoring the screen. These different roles were visible in the gaze behavior of the students of pair 2. We found that the sharing student's (2S) gaze behavior was more scattered compared with the monitoring student's (2M) gaze behavior during the physics content–related talk (Figures 8B, 9B). This behavior is logical because 2S had to divide their visual attention between multiple objects in the simulation view while controlling everything on the screen. Respectively, during the explication level, 2M was able to monitor the electric field by focusing their visual attention on the sensor (Figure 9B). From the perspective of these different roles, gaze dissimilarities between the students seem inevitable, emphasizing the consideration of contextual information on CKC processes when interpreting eye-tracking data and analysis (Liu et al., 2021). At best, the gaze dissimilarities between students could trigger critical discussion of the presented ideas and lead to higher CKC levels and improved learning outcomes.

Our study has certain limitations, such as using only two student pairs to illustrate our approach. This limit could be overcome in the future because the ENA scores (the summary statistic of the corresponding network) could be used to study the similarities and differences in the pairs' gaze behavior with larger sample sizes (for more details, see Shaffer et al., 2016; Andrist et al., 2018). In these cases, it is important to include only the necessary AOIs (the nodes of the epistemic networks) into the analysis so that the epistemic networks are easily interpretable. Moreover, the eye-tracking data analysis has some limitations when the data are collected in authentic, uncontrollable settings, as in our case: for example, the students were able to move freely during the data collection, so their visual scene was constantly changing when they moved their head and moved the objects in the simulation view. In our study, we aimed to improve the validity and reliability of the interpretations by analyzing the video data from three perspectives: the quantity of the talk, the quality of the physics content–related talk, and the temporality of CKC processes; and then exploring how eye-tracking data and analysis can shed light on CKC processes in terms of these three perspectives.

Despite these limitations, our study has several implications for future research. We illustrated how gaze behavior reflects the overall progress of CKC processes (Sections "Temporality of CKC processes" and "Gaze behavior sheds light on the temporality of CKC processes") and the different CKC levels (Sections "Quality of physics content–related talk" and "Gaze behavior sheds light on the knowledge construction approaches"). In particular, gaze behavior could be used to capture the different activities that the pairs (or groups, in general) conduct within a specific CKC level, during non-content-related talk, or during silent moments. For example, even though the students were silent even over half of the time (as was the case with pair 1), their gaze behavior during these silent moments may help us understand their success or failures in the CKC process. In our study, the gaze behavior of pair 1 indicated that they used these silent moments for writing their answers in the textbox, even though they had not made proper observations of the properties of the electric field. The gaze behavior of pair 2 indicated that they also used these silent moments for working with the simulation; this behavior might contribute to their iterative CKC process, in which they moved back and forth between the CKC levels and non-content-related talk.

As a methodological implication, we followed and extended Andrist et al.'s (2018) work by applying ENA to study gaze behavior in an authentic simulation-based environment. Our approach considered the spatial and temporal dimensions of the eye-tracking data, both of which provided essential information about the CKC processes. Our approach complements (instead of compensating for) the cross-recurrence quantification analysis, in which the focus is on the temporal alignment of students' gazes, here without spatial information about their visual attention. Thus, our study provides a novel approach for exploring CKC processes by combining video data and both spatial and temporal information from eye-tracking data. In the future, these explorations, together with learning outcomes, should be further investigated with larger sample sizes and in more diverse contexts. Future studies should also focus not only on the kinematics, but also the dynamics of these constructs, hence examining whether and why similar gaze behavior can lead to dissimilarities in the CKC process and its quality. Visualizations of CKC processes and gaze behavior could help the teachers and developers of educational technology design, implement, and refine productive CKC processes in simulation-based environments with appropriate forms of guidance. Contrary to the mobile, wearable eye trackers that we used in the current study, screen-based eye trackers could ease eye-tracking data analysis and visualization in computer-supported settings. Through the understanding of gaze behavior, one could envision a future where teachers use such trackers to guide and synchronize students' gaze in real time. Therefore, it is crucial to involve teachers and students in co-designing these visualization tools to increase their usability, transparency, and acceptability among practitioners (Buckingham Shum et al., 2019).

Conclusion

Typically, eye-tracking data analysis in CKC settings has focused on whether students are looking at the same objects but has done so without analyzing whether these objects are relevant to the problem at hand. We have illustrated how gaze behavior can shed light on CKC regarding the quantity of the talk, quality of the physics content–related talk, and temporality of CKC processes. These kinds of approaches may help teachers, researchers, and developers of educational technologies understand and guide CKC processes by showing the critical moments in these processes and revealing the features in the simulation environment that attract unnecessary visual attention. In the future, which kind of gaze-based indicators appropriately reflect the temporality of CKC processes and complement cross-recurrence quantification analyses should be considered. For example, when the CKC processes have low quality or move in the wrong direction, gaze dissimilarities could trigger critical discussion of the presented ideas and lead to higher CKC levels and better learning outcomes. Therefore, gaze dissimilarity can occasionally be essential for rising to higher-level CKC processes and for favorably advancing the solutions to a given problem.

Data availability statement

The datasets presented in this article are not readily available because video and eye-tracking data analyzed cannot be public due to personal data protection. More information on the data can be requested from the corresponding author. Requests to access the datasets should be directed to joni.lamsa@oulu.fi.

Ethics statement

The study was reviewed and approved by the Human Sciences Ethics Committee of the University of Jyväskylä. The participants provided their written informed consent to participate in this study.

Author contributions

JL: conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft, plus review, and editing), and visualization. JiK: conceptualization, investigation, data curation, and writing (review and editing). AL, PK, and TM: conceptualization and writing (review and editing). JaK: conceptualization, investigation, and writing (review and editing). RH: funding acquisition, conceptualization, and writing (review and editing). All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Academy of Finland (grant number 318905, the Multidisciplinary Research on Learning and Teaching profiles II of the University of Jyväskylä).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Andrist, S., Ruis, A. R., and Shaffer, D. W. (2018). A network analytic approach to gaze coordination during a collaborative task. *Comput. Hum. Behav.* 89, 339–348. doi: 10.1016/j.chb.2018.07.017

Becker, S., Mukhametov, S., Pawels, P., and Kuhn, J. (2021). "Using mobile eye tracking to capture joint visual attention in collaborative experimentation," in *Physics Education Research Conference 2021 Proceedings*. (eds.) M. Bennett, B. Frank, and R. Vieyra; College Park, US: American Association of Physics Teachers, 39–44.

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., et al. (2021). "The mathematical foundations of epistemic network analysis," in Proceedings of the Advances in Quantitative Ethnography: Second International Conference—ICQE 2020, eds. A. R. Ruis and S. B. Lee; Switzerland: Springer, 91–105

Buckingham Shum, S., Ferguson, R., and Martinez-Maldonado, R. (2019). Human-centerd learning analytics. *J. Learn. Analy.* 6, 1–9. doi: 10.18608/jla.2019.62.1

Chang, C.-J., Chang, M.-H., Liu, C.-C., Chiu, B.-C., Fan Chiang, S.-H., Wen, C.-T., et al. (2017). An analysis of collaborative problem-solving activities mediated by individual-based and collaborative computer simulations. *J. Comput. Assist. Learn.* 33, 649–662. doi: 10.1111/jcal.12208

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., and Fischer, F. (2020). Simulation-based learning in higher education: a meta-analysis. *Rev. Educ. Res.* 90, 499–541. doi: 10.3102/0034654320933544

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037/1040-3590.6.4.284

Clark, D., Nelson, B., Sengupta, P., and D'Angelo, C. (2009). *Rethinking Science Learning Through Digital Games and Simulations: Genres, Examples, and Evidence*. Washington, DC: Learning Science: Computer Games, Simulations, and Education Workshop Sponsored by the National Academy of Sciences.

Coco, M. I., and Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Front. Psychol.* 5:510. doi: 10.3389/fpsyg.2014.00510

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., and Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad. Med.* 88, 872–883. doi: 10.1097/ACM.0b013e31828ffdcf

de Jong, T., and Lazonder, A. W. (2014). "The guided discovery learning principle in multimedia learning," in *The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer, vol. *1* (Cambridge: Cambridge University Press), 371–390.

de Jong, T., Linn, M. C., and Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science* 340, 305–308. doi: 10.1126/science.1230579

Eagan, B., Brohinsky, J., Wang, J., and Shaffer, D.W. (2020). "Testing the reliability of inter-rater reliability," in *Proceedings of the Tenth International Conference on Learning Analytics and Knowledge* (LAK '20) (New York, USA: Association for Computing Machinery), 454–461

Hahn, L., and Klein, P. (2022). Eye tracking in physics education research: a systematic literature review. *Phys. Rev. Phys. Educ. Res.* 18:013102. doi: 10.1103/PhysRevPhysEducRes.18.013102

Hayashi, Y. (2020). Gaze awareness and metacognitive suggestions by a pedagogical conversational agent: an experimental investigation on interventions to support collaborative learning process and performance. *Int. J. Comput.-Support. Collab. Learn.* 15, 469–498. doi: 10.1007/s11412-020-09333-3

Hayashi, Y., and Shimojo, S. (2021). "Investigating gaze behavior of dyads in a collaborative explanation task using a concept map: Influence of facilitation prompts on perspective taking," in *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning—CSCL 2021*. eds. C.E. Hmelo-Silver, B. De Wever, and J. Oshima (Bochum, Germany: International Society of the Learning Sciences), 149–152.

Henri, F. (1992). "Computer conferencing and content analysis," in *Collaborative Learning Through Computer Conferencing. The Najadan Papers*. ed. A. R. Kaye (Berlin, Heidelberg: Springer-Verlag), 117–136.

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon. Bull. Rev.* 27, 856–881. doi: 10.3758/s13423-020-01715-w

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press

Jensen, J. L., and Lawson, A. (2011). Effects of collaboration and inquiry on reasoning and achievement in biology. *CBE Life Sci. Educ.* 10, 64–73. doi: 10.1187/cbe.10-07-0089

Jeong, H., and Hmelo-Silver, C. E. (2010). Productive use of learning resources in an online problem-based learning environment. *Comput. Hum. Behav.* 26, 84–99. doi: 10.1016/j.chb.2009.08.001

Jeong, H., Hmelo-Silver, C. E., and Jo, K. (2019). Ten years of computer-supported collaborative learning: a meta-analysis of CSCL in STEM education during 2005–2014. *Educ. Res. Rev.* 28:100284. doi: 10.1016/j.edurev.2019.100284

Jeong, H., Hmelo-Silver, C., and Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *Int. J. Comput.-Support. Collab. Learn.* 9, 305–334. doi: 10.1007/s11412-014-9198-3

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037/0033-295X.87.4.329

Klein, P., Viiri, J., Mozaffari, S., Dengel, A., and Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: how do students look at vector field plots? *Phys. Rev. Phys. Educ. Res.* 14:010116. doi: 10.1103/PhysRevPhysEducRes.14.010116

Lämsä, J., Hämäläinen, R., Koskinen, P., and Viiri, J. (2018). Visualising the temporal aspects of collaborative inquiry-based learning processes in technology-enhanced physics learning. *Int. J. Sci. Educ.* 40, 1697–1717. doi: 10.1080/09500693.2018.1506594

Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., and Lampi, E. (2021a). What do we do when we analyse the temporal aspects of computer-supported collaborative learning? A systematic literature review. *Educ. Res. Rev.* 33:100387. doi: 10.1016/j.edurev.2021.100387

Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., and Mannonen, J. (2020). The potential of temporal analysis: combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Comput. Educ.* 143:103674. doi: 10.1016/j.compedu.2019.103674

Lämsä, J., Uribe, P., Jiménez, A., Caballero, D., Hämäläinen, R., and Araya, R. (2021b). Deep networks for collaboration analytics: promoting automatic analysis of face-to-face interaction in the context of inquiry-based learning. *J. Learn. Analy.* 8, 113–125. doi: 10.18608/jla.2021.7118

Lehtinen, A., and Viiri, J. (2017). Guidance provided by teacher and simulation for inquiry-based learning: a case study. *J. Sci. Educ. Technol.* 26, 193–206. doi: 10.1007/s10956-016-9672-y

Lin, T.-J., Duh, H. B.-L., Li, N., Wang, H.-Y., and Tsai, C.-C. (2013). An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. *Comput. Educ.* 68, 314–321. doi: 10.1016/j.compedu.2013.05.011

Liu, C. C., Hsieh, I. C., Wen, C. T., Chang, M. H., Fan Chiang, S. H., Tsai, M.-J., et al. (2021). The affordances and limitations of collaborative science simulations: the analysis from multiple evidences. *Comput. Educ.* 160:104029. doi: 10.1016/j.compedu.2020.104029

Marquart, C.L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., and Shaffer, D.W. (2021). rENA: epistemic network analysis (0.2.3). Available at: https://CRAN.R-project.org/package=rENA

Mavin, T. J., Kikkawa, Y., and Billett, S. (2018). Key contributing factors to learning through debriefings: commercial aviation pilots' perspectives. *Int. J. Train. Res.* 16, 122–144. doi: 10.1080/14480220.2018.1501906

Molenaar, I. (2021). "Personalisation of learning: towards hybrid human-AI learning technologies," in *OECD Digital Education Outlook 2021: Pushing the Frontiers With Artificial Intelligence, Blockchains and Robots*. ed. S. Vincent-Lancrin, vol. *1* (Paris: OECD Publishing), 57–78.

Olsen, J. K., Sharma, K., Rummel, N., and Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *Br. J. Educ. Technol.* 51, 1527–1547. doi: 10.1111/bjet.12982

Richardson, D. C., and Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cogn. Sci.* 29, 1045–1060. doi: 10.1207/s15516709cog0000_29

Rutten, N., van Joolingen, W. R., and van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Comput. Educ.* 58, 136–153. doi: 10.1016/j.compedu.2011.07.017

Schellens, T., and Valcke, M. (2005). Collaborative learning in asynchronous discussion groups: what about the impact on cognitive processing? *Comput. Hum. Behav.* 21, 957–975. doi: 10.1016/j.chb.2004.02.025

Schneider, B. (2019). "Unpacking collaborative learning processes during hands-on activities using mobile eye-trackers," in *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL)*. eds. K. Lund, G.P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, and M. Baker (Lyon, France: International Society of the Learning Sciences), Vol. 1, 41–48.

Schneider, B., Dowell, N., and Thompson, K. (2021). Collaboration analytics—current state and potential futures. *J. Learn. Analy.* 8, 1–12. doi: 10.18608/jla.2021.7447

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., and Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *Int. J. Comput.-Support. Collab. Learn.* 13, 241–261. doi: 10.1007/s11412-018-9281-2

Schotter, E. R., Angele, B., and Rayner, K. (2012). Parafoveal processing in reading. *Atten. Percept. Psychophysiol.* 74, 5–35. doi: 10.3758/s13414-011-0219-2

Shaffer, D.W. (2017). *Quantitative Ethnography*. Madison, Wisconsin: Cathcart Press

Shaffer, D. W., Collier, W., and Ruis, A. R. (2016). A tutorial on epistemic network analysis: analyzing the structure of connections in cognitive, social, and interaction data. *J. Learn. Analy.* 3, 9–45. doi: 10.18608/jla.2016.33.3

Sharma, K., Olsen, J.K., Verma, H., Caballero, D., and Jermann, P. (2021). "Challenging joint visual attention as a proxy for collaborative performance," in *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning—CSCL 2021*, eds. C.E. Hmelo-Silver, B. De Wever, and J. Oshima (Bochum, Germany: International Society of the Learning Sciences), 91–98.

Stevenson, M., Lai, J. W. M., and Bower, M. (2022). Investigating the pedagogies of screen-sharing in contemporary learning environments—a mixed methods analysis. *J. Comput. Assist. Learn.* 38, 770–783. doi: 10.1111/jcal.12647

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., and Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: a systematic literature review. *Educ. Stud. Math.* 104, 147–200. doi: 10.1007/s10649-020-09948-1

Tatler, B. W., Kirtley, C., MacDonald, R. G., Mitchell, K. M. A., and Savage, S. W. (2014). "The active eye: perspectives on eye movement research" in *Current Trends in Eye Tracking Research*. eds. M. Horsley, M. Eliot, B. A. Knight and R. Reilly (Switzerland: Springer International Publishing), 3–16.

Tobii Pro AB (2017). Eye tracker data quality report: Accuracy, precision and detected gaze under optimal conditions—Controlled environment: Tobii Pro glasses 2 firmware v1.61. Available at: https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-glasses-2-accuracy-and-precision-test-report.pdf

Tobii Pro AB (2022). Pro Lab user manual v1.181. Available at: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/Tobii-Pro-Lab-User-Manual/

UC Berkeley (2022). Web-based inquiry science environment. WISE. Available at: https://wise.berkeley.edu/

University of Colorado Boulder (2022a). Charges and fields PhET simulation. Available at: http://phet.colorado.edu/sims/html/charges-and-fields/latest/charges-and-fields_en.html

University of Colorado Boulder (2022b). PhET interactive simulations. Available at: https://phet.colorado.edu/

Veerman, A., and Veldhuis-Diermanse, E. (2001). "Collaborative learning through computer-mediated communication in academic education," in *Proceedings of the Euro CSCL 2001 Conference*. eds. P. Dillenbourg, A. Eurelings, and K. Hakkarainen; Maastricht, The Netherlands: University of Maastricht, 625–632.

Yang, X., Li, J., and Xing, B. (2018). Behavioral patterns of knowledge construction in online cooperative translation activities. *Internet High. Educ.* 36, 13–21. doi: 10.1016/j.iheduc.2017.08.003

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/feduc.2022.1051098

Check for updates

OPEN ACCESS

EDITED BY
Muhammet Usak,
Kazan Federal University, Russia

REVIEWED BY
Milan Kubiatko,
J. E. Purkyne University, Czechia
Vincentas Lamanauskas,
Vilnius University, Lithuania

*CORRESPONDENCE
Martin Rusek
martin.rusek@pedf.cuni.cz

SPECIALTY SECTION
This article was submitted to STEM Education, a section of the journal Frontiers in Education

RECEIVED 22 September 2022
ACCEPTED 25 October 2022
PUBLISHED 16 November 2022

CITATION
Tóthová M and Rusek M (2022) "Do you just have to know that?" Novice and experts' procedure when solving science problem tasks.
Front. Educ. 7:1051098.
doi: 10.3389/feduc.2022.1051098

"Do you just have to know that?" Novice and experts' procedure when solving science problem tasks

Martina Tóthová and Martin Rusek*

Department of Chemistry and Chemistry Education, Faculty of Education, Charles University, Prague, Czechia

Only teachers who possess problem-solving skills can develop them in their students. These skills therefore need to be accentuated during teachers' pre-service training. In this study, attention was given to pre-service chemistry teachers' (students) problem-solving skills measured with the use of two sets of problem tasks–chemistry and general science tasks. Based on a pre-test consisting of both types of tasks, one successful, one partially successful and one unsuccessful solver was selected from a group of first-year bachelor chemistry teacher students. To compare, the tasks were also given to three experts (post-docs in the field of chemistry education). All the participants solved two tasks on a computer with their eye movements recorded. After the procedure, retrospective think-aloud and interviews were conducted to provide data about the problem-solving process. The results showed several trends. (1) Students–novices considered the chemistry task more difficult than the science task, which correlated with their task results. (2) Experts considered the science task more complex, therefore more difficult, however scored better than the students. (3) Even the successful student only solved the chemistry task using memorized facts without the support provided. (4) Experts' direct focus on relevant parts was confirmed, whereas unsuccessful (novice) students distributed their focus toward other task parts too. (5) When students faced a problem during task solving, they used limiting strategies. This behavior was not identified in the expert group. The results thus showed a need to support students' problem-solving strategies in several areas, especially careful reading, and identifying the main problem and supporting information. Moreover, the results showed a need to present chemistry tasks to students with more variability and explain their reasoning rather than testing field-specific, separated, memorized information.

KEYWORDS

problem-solving skills, chemistry education, science education, eye-tracking, pre-service teachers

Introduction

Problem-solving ability has been seen as crucial (Bellanca, 2010; OECD, 2016; De Wever et al., 2021) and is predicted to be one of the most important skills for the future (OECD, 2018). It is also strongly linked with the Program of International Student Assessment (PISA) research results which are widely discussed topic in the field of education as PISA is considered the most important international research in the field of measuring educational outcomes (Potužníková et al., 2014). However, it is often the target of criticism for possible errors or imperfections that may be caused by the measurement load or incorrect interpretation and uncritical approach to their results (Straková, 2016; Zhao, 2020). The general public, as well as the politicians of given countries, react to the test results. They are used to respond to current changes in educational policy (Gorur, 2016; Vega Gil et al., 2016). However, the indicator they are guided by is often only the ranking of the state's pupils (Kreiner and Christensen, 2014; Štech, 2018). Test scores alone may not be sufficient in the future, given that tests fail to capture subtle but important differences between students. New methods are required for teachers to determine whether students have truly understood a given topic (Tai et al., 2006).

Program of international student assessment (PISA) results consistently indicate that Czech students' ability in the area of problem-solving has been declining (OECD, 2016, 2019). The reasons behind these results remain hidden. The possible cause may be lack of its development in schools. Teachers play a crucial role in these skills' development see e.g., Tóthová and Rusek (2021b). Apart from including problem-tasks in their lessons (Lee et al., 2000), the teachers' own ability to solve problems is seen as necessary (Krulik and Rudnick, 1982). Only quantitative research and mere tests are not sufficient enough to get more information (e.g., Barba and Rubba, 1992). Correct answers in these tests may not refer to conceptual understanding of the problem (Phelps, 1996; Tai et al., 2006; Rusek et al., 2019). Some research aimed at this area used qualitative research (e.g., Cheung, 2009; Barham, 2020). More studies are needed to understand the state of these abilities in pre-service teachers and design university courses.

To elucidate the process of problem-solving, think-aloud (Rusek et al., 2019) and eye-tracking (Tsai et al., 2012) proved to be sufficient methods. With a combination of these methods, students' procedures in solving tasks can be described in more detail see e.g., Tóthová et al. (2021). A more in-depth analysis and investigation into the reasons for a student's choice of a given answer can be an indicator to further develop the monitored competences.

The present study therefore included all the above-mentioned aspects: used a combination of qualitative methods to evaluate pre-service teachers' problem-solving skills and processes. The results can bring the information needed to develop university courses as well as improving pre-service teacher training practice.

Theoretical background

This study was based on eye-tracking as with its expansion in education research see e.g., Lai et al. (2013), its use in analyzing science problem-solving has been increasing. Despite the possibilities such as eye-tracking goggles [used e.g., in research in laboratory (van der Graaf et al., 2020)], the use of eye-tracking methods in science education research still remains mainly in laboratory conditions with the use of screen-based or stationary eye-trackers (Tóthová and Rusek, 2021c).

When analyzing problem-solving with the use of eye-tracking, attention was paid to the use of scientific representations (Lindner et al., 2017; Klein et al., 2018; Wu and Liu, 2021). To analyze those, the specific metrics, such as number of fixations (NOF) or total fixation duration (TFD), proved to be useful. The results showed that the use of representations helped students understand the text and solve given problems (Lindner et al., 2017). In the same research area, Rodemer et al. (2020) explained the influence of previous knowledge and skills when using representations. More experienced students were able to make more transitions between the provided representations, therefore using the support they were given by the task itself. The use of these representations may be influenced by the strategies students apply (Klein et al., 2018) and also vice versa, the use of text and visualizations can influence the strategies used (Schnotz et al., 2014).

It is, therefore, not surprising that the strategies used during problem-solving are often the subject of research. The differences between novices and experts in the used strategies were shown in the study by Topczewski et al. (2017). With the use of order of fixations, they identified that novices approach the same problem using different gaze patterns, i.e., different strategies. The used strategies may be influenced by the type of instructions given to the students.

As the used strategies seem to be the dealbreaker in the problem-solving process, further understanding their influence is needed. Tsai et al. (2012) analyzed the difference between successful and unsuccessful solvers in multiple-choice science problem-solving. Their results showed that the differences lied in the successful solvers' focus on relevant factors, whereas the unsuccessful showed a higher frequency of observing irrelevant factors. In a similarly targeted study, (Tóthová et al., 2021) identified the reasons for unsuccessful problem-solving in chemistry on an example of their periodic table use. Apart from focusing on irrelevant factors, the use of limiting strategies and problems with reading was identified as the main problems. This rationale stood as the basis for the presented study, whose aim was to investigate the validity of some of the aforementioned findings in another context with the use of eye-tracking-based methods.

Aims and materials and methods

Aims

The aim of this study was to map pre-service chemistry teachers' procedures when solving problem tasks and to compare it with experts' procedures. The study followed these research questions:

I. What, if any, is the difference between the pre-service teachers' performance on chemistry and general science problem tasks?
II. What, if any, is the difference between pre-service teachers' and expert procedure when solving problem tasks?

Research procedure and participants' selection

The study was based on mixed methods. It used a pre-experimental design (**Figure 1**). At the beginning of the academic year 2021/2022, first-year bachelor university students who chose chemistry teaching as one of their majors at Charles University, Faculty of Education were given two sets of tasks focusing on chemistry and general science (see above). Based on their results, one successful, one partially successful, and one unsuccessful student were chosen for the follow-up eye-tracking study.

All the study participants were explained the purpose of the study and were ensured their results as well as any other data (eye-tracking and audio recordings) will remain available only to the researchers and will not be shared with a third party. For the purpose of the data presentation, the students and experts were anonymized. All the data were saved on one computer protected by a password.

It consisted of them solving another set of problem tasks of comparable difficulty (see below) with their eye-behavior being recorded by an eye-tracking camera. After calibration, participants were asked to sit still and watch the screen. The time for solving the task was not limited. Another task occurred to the solvers after a random key click. After they finished the last task, the cued retrospective think-aloud method (Van Gog et al., 2005) was used to understand the participants' mental processes. Based on the students' explanations of their problem-solving processes, interviews were conducted.

As the think-aloud method is based on the solvers' explanation of their solving process and a researcher should intervene only when a respondent stops talking (Van Someren et al., 1994), the semi-structured interviews conducted after RTA provide even more information. The follow-up interview topics aimed at: task difficulty (perceived difficulty, most difficult part, more difficult task); solving confidence; type of task (similarity and differences between given tasks, similarity and differences between given task and tasks they were used to). Students attended the experiment voluntarily. During the eye-tracking measurements, they sat approximately 70 cm from the computer screen. Respondents' think-alouds and interviews were recorded. The entire eye-tracking phase took approximately 45 min.

To compare the task-solving process, three experts on chemistry education (all owning a Ph.D. in the field) were chosen from the researchers' department's academic staff and participated in the eye-tracking, RTA, and interviews.

Research tool

Two sets of complex tasks[1] were prepared: one for the pre-test and one for the post-test. The pre-test tasks consisted of three complex chemistry problem tasks and three complex general scientific literacy problem tasks. The test used in the eye-tracking research phase consisted only of one complex task from each group "a chemistry task" and "a PISA task," both originally designed for 15-year-old students. The tasks focused on working with available information and visualized data. Both chemistry tasks were taken from the Czech national chemistry curriculum standards (Holec and Rusek, 2016) from which tasks focused on general chemistry were chosen. Both the PISA-like scientific literacy tasks were represented by selected items released from PISA (Program of International Student Assessment) task pilots (Mandíková and Houfková, 2012). These tasks focused on students' ability to plan and evaluate scientific research as well as their ability to gather information from diagrams, tables, etc.

To ensure the pre- and eye-tracking test tasks were on the same level of difficulty, "optimum level" chemistry tasks and the same scientific literacy level tasks were chosen. Also, the authors' research group members evaluated the tasks to prevent one set from being more difficult than the other.

In the chemistry task, the solvers were given three sub-tasks (referred to as Task 1–3). The first dealt with the trends in the periodic table (halogens) and their solution required working with the periodic table and using information from a text. The second subtask targeted the reaction rate of alkali metals with water. A description of lithium and sodium's reaction with water was given and the solvers were asked to select correct

1 Complex tasks are tasks consisting of several subtasks with a different focus connected by a unifying topic.

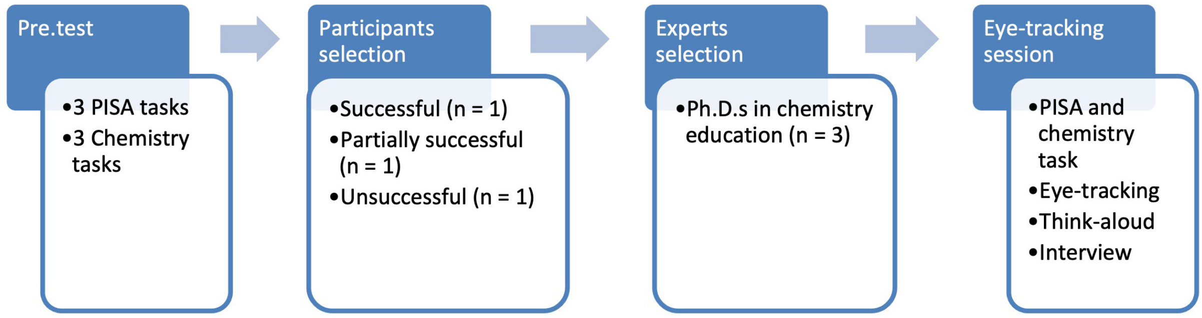


FIGURE 1
The study design.

descriptions of potassium's reaction with water. In the last subtask, a general description of the trends in the periodic table regarding atomic diameter was given and the solvers were required to order the given elements based on it.

The PISA task was set into a classroom problem–students want to design a nearby river's pollution measurements from physical, chemical and biological points of view. There were four subtasks (also referred to as Task 1–4) to this complex task. The first asked about important and unimportant equipment, in the second, the students were supposed to describe a process of water-flow measurement. The third was about an appropriate sampling site selection on a map based on concrete criteria and the last about pollution data evaluation based on data in a table.

Apparatus

Tobii Fusion Pro with a 250 Hz sampling rate and GazePoint eye-camera GP3 (60 Hz) were used in this study. Prior to all recordings, nine-point calibrations were performed. Both instruments allowed respondents' free head movements.

Data analysis

To evaluate the PISA-like tasks, the scoring used in the original research was taken over (Mandíková and Houfková, 2012; OECD, 2016), i.e., 2 points for a completely correct result, 1 point for partially correct and 0 points for an incorrect result. The chemistry tasks were evaluated in the same way. To ensure objectivity, two researchers evaluated the tests independently and compared. The scores were reported as a percentage for later comparison.

To analyze participants' task-solving procedure, time fixation duration (TFD) see e.g., Lai et al. (2013) in pre-selected areas of interest (AOIs) was measured. The data are reported as proportion of total fixation duration and fixation duration in particular AOI, i.e., proportion of time fixation duration see e.g., Jian (2018). The AOIs included the task itself, answer choices area, visual parts and contexts or information. The data were analyzed with the use of IMB SPSS completed with the retrospective-think aloud recordings, as well as interviews. Both were recorded, transcribed verbatim and coded using the set of codes for students' strategies taken from previous studies (Rusek et al., 2019; Tóthová et al., 2021). They were then divided according to Ogilvie (2009)–expansive and limiting. The software QDA Miner Lite was used for transcription analysis.

The reported confidence values allowed to relate students' and experts' confidence with their answers (score). These data were analyzed with respect to Caleon and Subramaniam (2010). Standard metrics as CAQ (mean confidence accuracy quotient) and CB (confidence bias) were analyzed.

The Shapiro–Wilk normality test ($p > 0.05$) showed the data were normally distributed with an exception of two observed areas of interest ($p = 0.06$ resp. 0.08). For this reason, differences between the students and experts' TFD and t-test were used to analyze the majority of TFDs, Mann–Whitney test for the non-normally distributed data. Cohen's d resp. r were used as effect-size tests and were interpreted according to Richardson (2011).

Results

Overall results

The overall students and experts' results (**Figure 2**) reveal several findings surprising in the light of the chosen tasks' nature. The data in **Figure 1** need to be read with discretion as the total number of points was low and a loss of one point reflects dramatically on the test score.

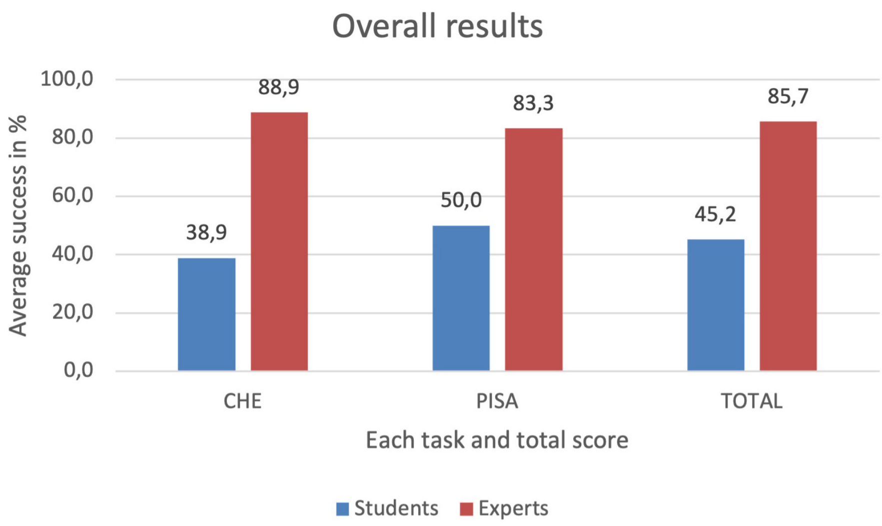


FIGURE 2
Participants' performance in program of international student assessment (PISA) and chemistry (CHE) problem tasks.

TABLE 1 Confidence related metrics.

Metric	Task/Group	CHE Task 1	CHE Task 2	CHE Task 3	PISA Task 1	PISA Task 2	PISA Task 3	PISA Task 4
CAQ	Students	−5.46	−1.36	−3.68	−1.60	−4.66	−0.93	−7.34
CB		−0.03	0.23	0.43	0.50	0.57	0.00	0.03
CAQ	Experts	0.97	0.99	0.87	0.00	0.98	0.99	−16.37
CB		−0.03	0.00	−0.13	0.33	−0.02	0.00	0.30

To answer the first research question, the students' and experts' results did not statistically significantly differ ($p = 0.149$) as expected in such a small sample. However, the effect size (d = 1.104) showed a large effect which points to the experts' significantly better performance. More information to answer the second research question is shown with the use of more tools: tasks, eye-tracking, retrospective think-aloud, and interviews.

The interviews showed the students overestimated their solution results. This was also reflected in their less accurate confidence scores (see **Table 1** showing confidence related metrics).

The experts' mean confidence accuracy quotient (CAQ) was positive compared to the students'. The differences were statistically significant with a large effect (p = 0.025; r = 0.598). The experts showed a higher ability to discern between what they know and do not know. On the contrary, the students' CAQ values suggest they tend to overestimate themselves and report higher confidence despite their task solution being incorrect. The confidence bias (CB) score confirmed students' tendency to overestimate themselves, whereas in four tasks, experts showed almost perfect to perfect calibration see Liampa et al. (2019). There was no statistically significant difference between students' and experts' CB (p = 0.123); however, the effect size (r = 0.413) showed a medium effect.

The eye-tracking results

The eye-tracking data served not only as a cue for the respondents' retrospective think aloud but also as a marker of the attention both the students and experts dedicated to particular parts of the tasks. As far as the chemistry task was concerned, the students solved it on average in 5 min as compared to the experts' 3 min 48 s. However, the unsuccessful student spent as much time on the chemistry task as the experts' average, which was not enough. On the contrary, the successful one spent as much time on the task as one of the experts (over 5 min), which paid off. As for the PISA task, the students spent over 8 min solving in comparison with the experts' 6 min 46 s. Again, the successful student took the longest to solve this task and the unsuccessful student used as much time as one of the experts.

This result suggests that the experts are efficient and need less time, the successful student worked their way to the solution and the unsuccessful student turned to a quick solution which did not lead to the desirable result. Naturally, this

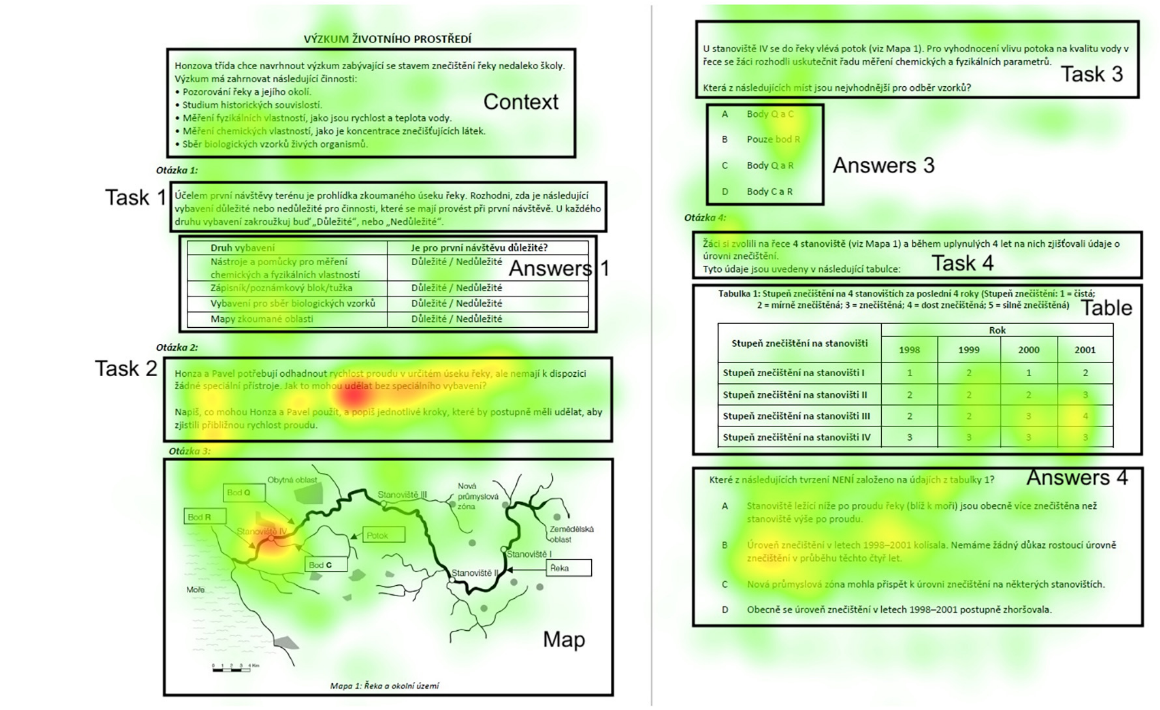


FIGURE 3
Experts' focus in the program of international student assessment (PISA) task.

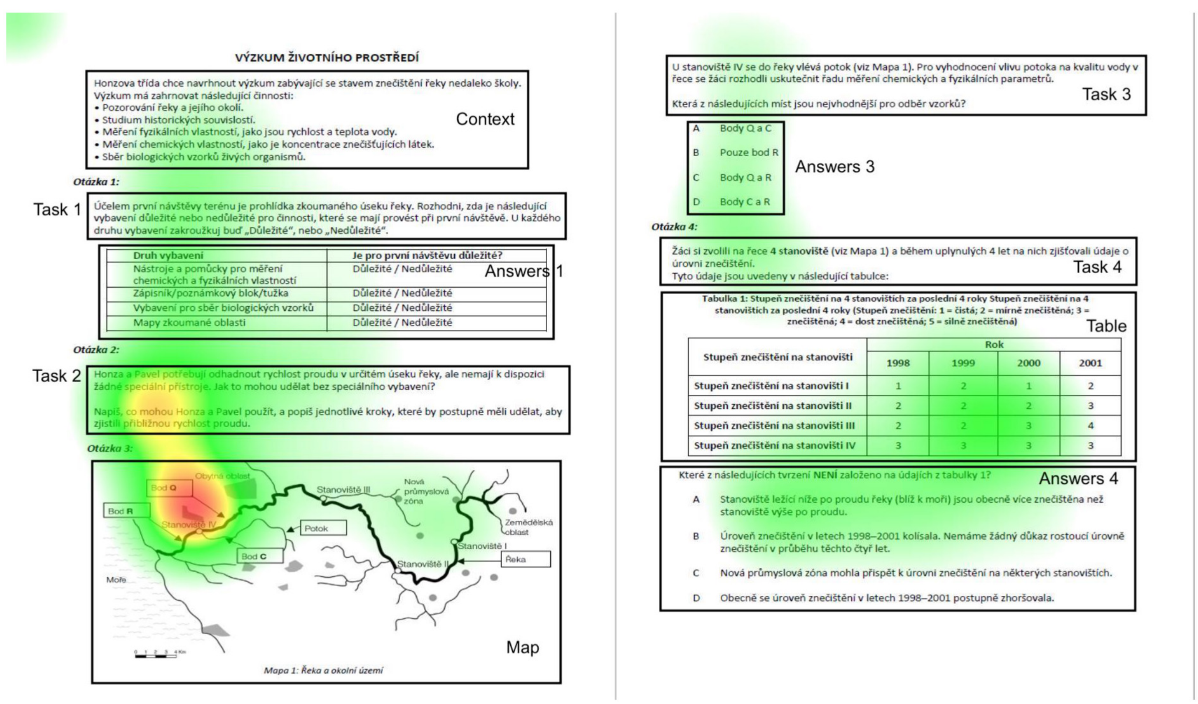


FIGURE 4
Students' focus in the program of international student assessment (PISA) task.

finding needs to be explained in more detail with the use of qualitative methods.

The program of international student assessment task

As far as this set of subtasks (Task 1–4) was concerned, no statistically significant differences were found on a 5% level of significance. This was expected due to the low sample size. However, when effect size was considered, five items showed a considerable power of effect-size. The experts spent more time on the task's context (see attention map on **Figure 3**). The effect size had a huge power (d = 1,96), which shows they put an effort into understanding what was requested from them.

The color gradient (from red through yellow and green to white) relates to the participants' attention and shows what part of the screen/task/area a solver focused on.

In the subtasks (Task 1–3), the students spent more time on Task 1–selecting useful equipment, and Task 4–choosing data from a table. The power of the effect was huge in both cases (d_{Task1} = 1.314, d_{Task4} = 1.645) suggesting the experts' greater efficiency as not only did they need less time but their answers were also correct.

On the contrary, on Tasks 2 and 3, experts used more time fixating the tasks with a huge (d_{Task2} = 1.651) and medium (d_{Task3} = 0.62) effect size, which shows they paid more attention to the tasks. In the case of Task 2, it was their careful reading in order not to miss anything from the task. In Task 3, they spent the same time on the given map of a river as the students, but more time on the text.

This action was then reflected not only in their correct results but also in the amount of time spent on Task 3's multiple-choice answer part (again the power of the effect was huge d = 1.293) which shows they just picked the correct response, in contrast to the students who spent more time on this part searching for the correct one (see **Figure 4**).

The chemistry task

As far as the chemistry task was concerned, surprisingly, neither the students nor the experts worked with the provided periodic table as much as was expected (see **Figures 5**, **6**).

One obvious difference appeared when their time fixation duration was observed on the table's legend (explaining the element groups). The students spent significantly more time on the legend. The difference was also statistically significant (p = 0.046), with a large effect size (r = 0.813) which shows the experts' expected familiarity. Another proof of their expertise was shown in Task 3's context. They needed significantly less time, although with just a medium effect size (d = 0.696), which showed their familiarity with the trends. On this very task, however, the large effect size (d = 0.849) showed experts spent significantly more time fixating the task. This proved to be necessary for the correct results, however, more reasoning was shown in the retrospective-think-aloud and interview (see below). A similar difference with a medium effect-size (r = 0.445) was found in Task 1, on which, again, experts spent more time than the students. Considering the task was not about simple recollection from memory, the eye-tracking data helped to reveal the reason for the experts' success. Nonetheless, it was the think-aloud and interviews which fully explained the reasons.

Retrospective think-aloud results: Applied strategies

The data above showed a difference between the students and experts' time fixation durations on different tasks. When the eye-behavior was replayed to each of the task solvers, their verbal description of the process helped identify the reasons behind their performance, i.e., use of strategies and facing problems. Although not all experts reached the maximum score in both (PISA and chemistry) tasks, their solving strategies differed from the students' considerably. The number of coded strategies was similar in both analyzed groups (69, resp. 70). For the sake of more accurate interpretation, absolute numbers and relative representations are reported in **Table 2**.

Both students and experts applied mostly expensive strategies such as e.g., *working with available information* (tables, map, and information in assignment). An example is given for chemistry Task 2 in which the solvers were supposed to infer on the course of reaction between alkali metals with water: *"Then I looked at the table, where I actually assessed that the order of the elements is lithium, sodium, potassium, according to which the reactivity due to the assignment should increase."* Surprisingly, the eye-tracking data did not show a difference in the time fixation duration, despite the experts being familiar with these reactions and not being expected to use inductive thought processes with the use of the periodic table.

Another example of an expansive strategy in use is reflection. In PISA Task 2, in which the solvers were supposed to choose appropriate parts of a river for water sampling, one of them mentioned *"…then I actually found out…, I thought that it would be appropriate to measure all three points < on the river > , and as soon as the answer wasn't there, I found out that the question is probably different than I initially understood…."*

One type of strategy was found only in experts' progress–mentioning alternative solutions. The example is from PISA Task 3: *"When they go to map it < the area in the task > , they have to write it down somewhere. But here I mentally came across the fact that when I want to record something, I usually take my phone. Or I turn on the navigation and don't need a map at all."* This strategy explains the same time spent on the map and also the longer time fixation duration on this task's text.

The difference between students and experts in using reading strategies was only slightly shown by eye-tracking–shorter time spent on certain texts–was skipping unnecessary

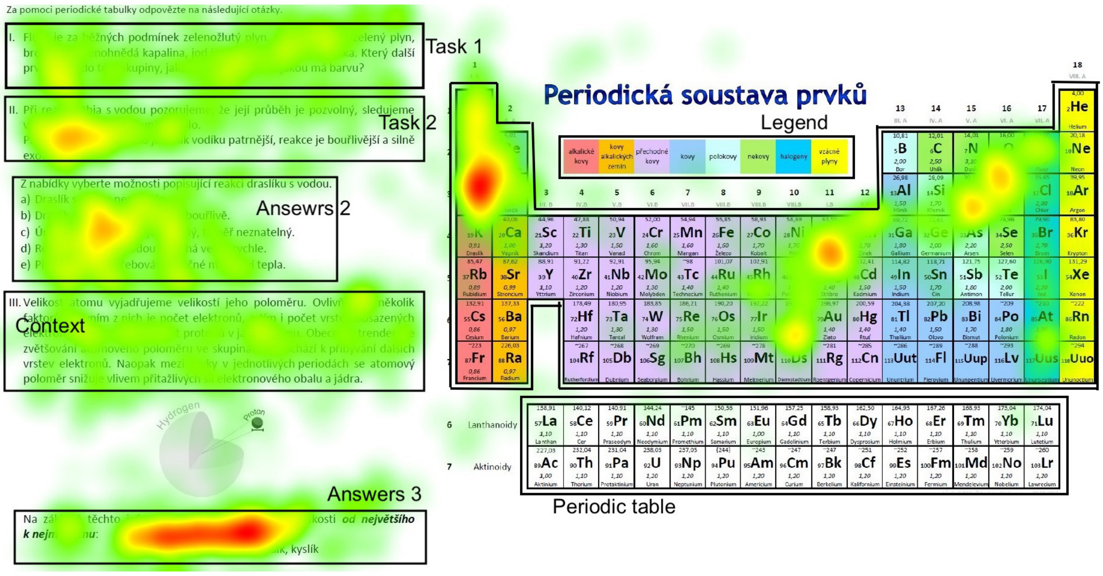


FIGURE 5
Experts' focus in the chemistry task.

information. This behavior was not identified during the think-aloud with the students: *"I skipped the picture because I was looking at what to do first, if I had needed it, I would have come back to it later."*

The possible key to the expertise in problem-solving is to avoid limiting strategies. In some cases, they may lead to the correct answer, but they are not applicable in any other cases and therefore represent unwanted behavior (Ogilvie, 2009). Students applied these strategies in almost 22% of the problem-solving process. Students, for example, used a guessing strategy when they faced a problem with a lack of knowledge. An example from chemistry Task 2: "*...the second task, I had no idea, so I read the options and guessed...*" This was possible to observe in the students' eye-movements (see **Figure 7**) compared with an expert (see **Figure 8**), who read the task and answer choices. The students focused mainly on the answer choices, which reflected their guessing strategy.

This answer offers an explanation for the experts and students' similar time fixation duration on this task. Whereas the experts' was shorter due to their familiarity with the reaction, the students skipped this task once they did not know the correct response without trying to figure it out. This behavior, which also appeared in other tasks, was later explained in the interviews–students are not used to solving complex problem tasks in chemistry. Therefore, whenever there is not an immediately obvious solution, they turn to a limiting strategy and/or give up.

On the contrary, the experts did not use limiting strategies even when they faced problems (see **Tables 2**, **3**). Here is a possible connection between their reported confidence. Students aware of guessing admitted it in the confidence scales; however, the experts did not have to. In order to promote students' problem-solving skills, it seems necessary to concentrate on the processes following the problem's identification.

Faced problems

As well as in the case of strategies, the problems students and experts faced differed. The results are shown in **Table 3**.

Students had problems with reasoning, e.g., in the case of PISA Task 2 concerning measuring the speed of the river: "*...I mentioned that it could be the bottle. At what speed the water will flow to the bottle...*" The problem with the lack of knowledge was surprising as the task aimed at basic knowledge in the respondents' study field and working with given information seemed natural. Nevertheless, the problem with applying knowledge also occurred in one of the experts' think-aloud comments. In chemistry Task 1, they were supposed to induce the color of an element based on a trend in the periodic table. The expert mentioned: "*...I used the table with the trends, but I did not know the color.*" This points to a finding also observed with the students–in spite of the correct mental process, this expert did not respond just because they did not recall the correct answer.

Another common problem was misunderstanding the assignment. This was related to the type of task (complex tasks with context). Students' answers were often contradictory to the assignment. The following example regarding chemistry

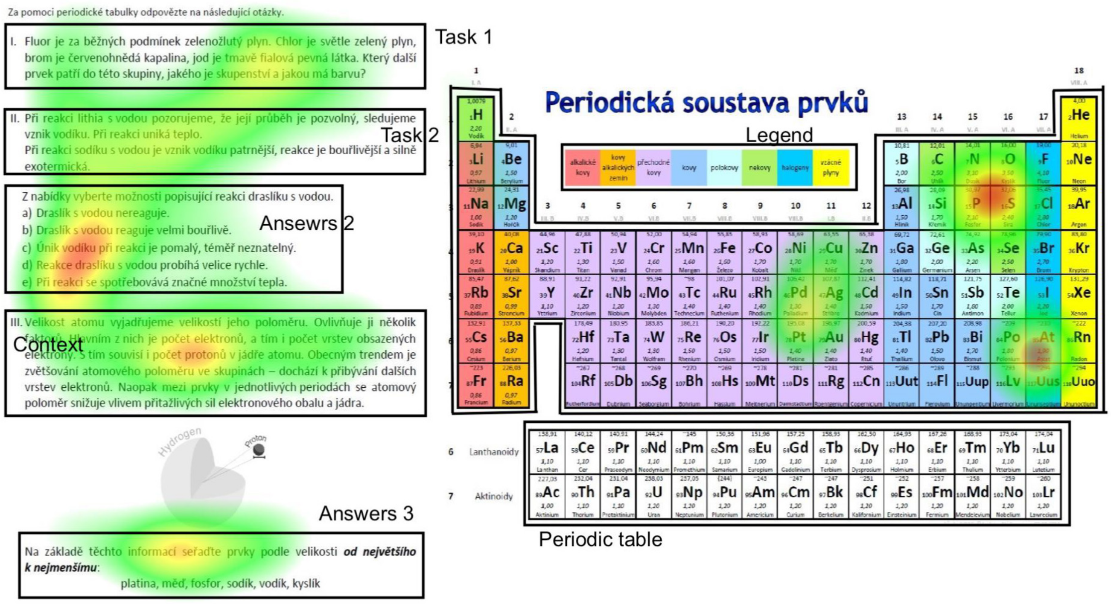


FIGURE 6
Students' focus in the chemistry task.

Task 3 on atomic diameter: "*...I usually read it twice, because the elements are smaller when they have a smaller number of electrons, protons, neutrons, so I found the elements and sorted them according to the proton number,*" in contrast to the information in the task itself (abbreviated): "*The trend is the increase of the atomic radius with increasing proton number in groups. Conversely, between elements in periods, the atomic radius decreases with increasing proton number...*"

Participations' evaluation of the tasks

The interviews served to help the researchers understand the observed thought processes manifested by eye-behavior and later verbalized during the think-aloud phase. They also provided information about the participants' evaluation of the tasks. The main answers from the interviews are shown in Table 4.

The interviews revealed a different attitude to the tasks. The students perceived the chemistry tasks as more difficult, due to the knowledge needed and the presence of the periodic table: "*Chemistry was more difficult–it is more theoretical and also, there was a table that confuses me....*" On the contrary, the experts assumed the PISA task would be more difficult for students because it included more data analysis, more logical reasoning and less information was given: "*Basically, I think PISA will be harder for them–it's completely up to the student. It doesn't have any supportive info, whereas there's always a hint in the chemistry one. PISA also depends more on reading–like table data....*"

Both students and experts mentioned orientation in the tables and text as the difficult part in solving the tasks. The above-mentioned attitude toward chemistry task solving was therewith discovered. Students, in accordance with their predecessors (Tóthová et al., 2021), reported the periodic table as something that confuses them: "*...if the table wasn't available*

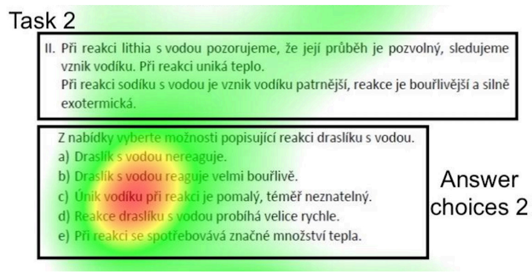


FIGURE 7
Student's attention to the answers' choices.

TABLE 2 Applied strategies.

	Strategies							
	All		**Expansive**		**Limiting**		**Reading**	
Students	69	100%	37	53.6%	15	21.7%	17	24.6%
Experts	70	100%	46	65.7%	0	0%	24	34.3%

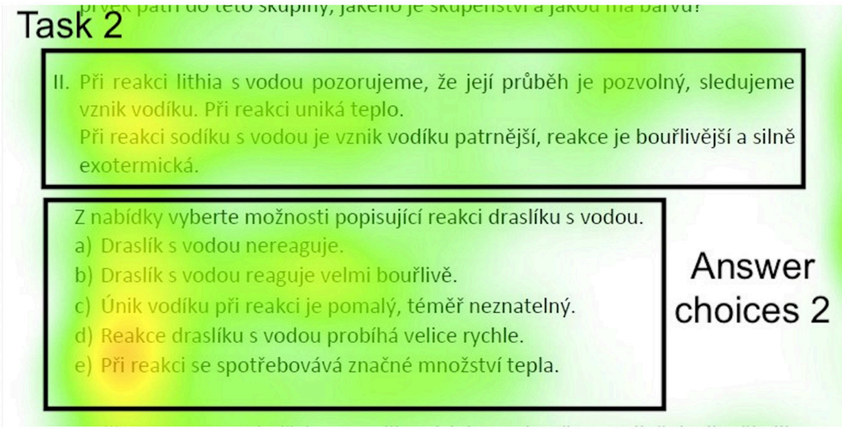


FIGURE 8
Expert's attention to the task itself and answer choices.

here, I'd be more confident. The presence of the table is a distracting element." As already pointed out in the cited research, the original idea of an inductive tool, which the periodic table undoubtedly, seems hidden even to pre-service teachers.

However, the students and experts differed in the perceived tasks' demands. The students saw the difference between the chemistry and PISA task mainly in the perceived needed knowledge in the chemistry task. They did not notice the purpose of the chemistry task (they were supposed to work with available information and periodic table). Only basic knowledge required in lower elementary school chemistry was needed. Another discovered difference between the students' and experts' opinion of the tasks resided in the theoretical basis of the chemistry task: *"The chemical one is more theoretical and the other practical."* However, this student opinion was in accordance with one expert: *"For the PISA task, they need to have experience with being outside in the scientific context. They will probably have one from school for the chemistry task."*

The results seem to reflect student-perceived nature of tasks and thus confirm the presumption formulated by Vojíø and Rusek (2022)–presenting students with a limited variation of tasks affects their abilities.

The interviews revealed the students were not used to solving these types of tasks. They agreed that tasks at school (of the "textbook genre") did not require logical reasoning and thinking or working with information, but rather memorized knowledge: *"The difficult thing here is that one has to think about what and how to answer, what makes sense and what doesn't. For example, in the tests we were given (at school), we would have to say what we memorized. So I would specifically answer the questions and I was sure of the answer there too."* Experts in the field of chemistry education who regularly visit lower and upper secondary schools answered similarly and added: *"Such tasks appear rather rarely."* One expert referred to one research's finding (Son and Kim, 2015) that *"when a more difficult task appears, they tend to divide it. At the same time, according to my experience, they do not work with tables and diagrams much. They are not even used to reading text."*

Discussion

Overall results

This study brought several results which inform not only the Czech education system but also the international science education community. First, the finding that pre-service teachers struggle with tasks developed for 15-year-olds suggests that upper-secondary school teachers consider success quite differently to the OECD (producing PISA tasks) and even national curriculum standards. In fact, this study confirmed previous studies' results showing deficiencies in students' achievement of curricular objectives (cf. e.g., Medková, 2013; Rusek and Tóthová, 2021).

The students with lower expertise (compared to experts) achieved worse results, which confirmed previous research results. For example, in a study from physics education (Milbourne and Wiebe, 2018), the importance of content knowledge was shown as one of the key factors affecting students' results. This was highlighted by the areas they paid attention to. Also, Harsh et al. (2019) found a significant difference in students' and experts' ability to read graphical data representations associated with the latter group's direct search patterns resulting in better results.

The worse student achievement in comparison with the experts was associated with their poorer ability to estimate the correctness of their own solution. This phenomenon is consistent with other research (e.g., Talsma et al., 2019; Osterhage, 2021).

TABLE 3 Faced problems.

		Number and type of problem	Number
Students	21	Non-logical reasoning	7
		Lack of knowledge	5
		Misunderstanding the assignment	5
		Redundant and incorrect information	4
Experts	3	Problem with the assignment	2
		Lack of knowledge	1

The differences between chemistry and program of international student assessment tasks (RQ1)

Student statements during the RTA, as well as interviews, helped understand this contradiction in more detail. The tasks presented to the students tested their literacy, science thinking and required a certain level of competency.

Despite the fact that respondents were chosen from pre-service chemistry teachers, students were less successful in the chemistry tasks than in general scientific tasks. This was proven even in a previous study in freshmen and students in the last year of their studies (Tóthová and Rusek, 2021a). The fact students saw chemistry tasks as more difficult than the PISA tasks and also the students' statements about the expected nature of chemistry tasks suggests students are used to a certain type of chemistry task. These are usually based on knowledge and do not require any higher order thinking such as analysis or synthesis of information from several sources. Even successful solvers confirmed they preferred to use memorized facts than the periodic table, which "confuses them." This was proven by the solvers' attention (reflected in TFD) paid to the periodic table, which was unexpectedly low. Thereby, a trend was confirmed pointing to a certain task culture for lower-secondary science textbooks (Bakken and Andersson-Bakken, 2021; Vojíø and Rusek, 2022). Also, the statements regarding the difference between the chemistry and PISA tasks which referred to the "need" of knowledge and memorized information when solving chemistry content tasks confirm the misunderstanding of chemistry's nature, resp. the periodic table (cf. Ben-Zvi and Genut, 1998). The interesting thing is that in the scientific (PISA) task, students did not mention needed knowledge; however, there were some scientific concepts, e.g., speed calculations.

On the other hand, experts considered the science task more complex, requiring more steps to solve them. Therefore, the PISA task was, according to them, more difficult.

The differences between students and experts (RQ2)

Differences were identified in the groups' eye-tracking records. TFD spent on defined AOIs alone cannot explain the solving process. Combining it with other methods was therefore necessary. Differences between students and experts were evident in the strategies used. Experts used only expansive strategies, whereas students tended to also use limiting strategies, e.g., guessing. The use of this strategy was found in previous research and was related more likely to the group of low achievers in the case of test-taking strategies (Stenlund et al., 2017). The use of limiting strategies also seemed to lead to incorrect solutions in this study. The reasons for using these strategies were the identified problems: lack of knowledge, non-logical reasoning or misunderstanding the assignment. This is consistent with previous research (Tóthová et al., 2021). However, when experts faced a problem, they did not use the limiting strategy and continued using expansive strategies. At

TABLE 4 Perceived task difficulty and differences.

Students	More difficult task	The most difficult part in solving	The difference CHE × PISA	The difference with task they are used to
1	Chemistry	Lack of information	In the CHE task, I simply have to know that.	At school, we had to write what we knew by heart.
2	Chemistry	Reading maps and tables	In the CHE task, knowledge is needed.	At school, we would write formulas which we learned before.
3	Chemistry	Reading and using table	The CHE one is more theoretical.	At school, it was just a question, not a comparison for example.
Experts				
1	PISA	Reading	You have to have experience when solving PISA one.	These tasks are far away from school reality, where teachers aim to remembering and understanding. In the presented tasks, there is designing of experiments, etc.
2	PISA	Logical reasoning	The CHE one is supported more with visual aid	In these tasks, the information was given, and students should work with them. At schools, teachers aimed at remembering facts.
3	PISA	Reading	The PISA one is more complex.	Teachers divide difficult tasks into less difficult ones. Also, reading is not developed.

the same time, this finding further stresses the importance of a more cautious approach to complex tasks such as those used in the PISA framework. Its robust task piloting as well as large samples on one hand limit external factors, nevertheless in light of this study's (and other) results, a number of false-positive results is very likely and merely dividing students according to literacy levels hardly provides sufficient information to suggest concrete changes in classroom instruction (cf. Tóthová and Rusek, 2021b).

Also, the problems faced by students and experts differed. The problem with misunderstanding the task leading to the incorrect solution occurred in students. Eye-tracking enabled these problems to be connected to the attention paid to concrete areas in the tasks. The difference between attention paid to concrete AOIs in students and experts occurred in the context part. Whereas in the PISA task, which dealt with general scientific knowledge, experts spent more time reading the task context, in the field of their study (chemistry) their spent significantly less time on the task context. This may reflect experts' ability to choose relevant information for their solution (c.f. Tsai et al., 2012), whereas students still have to learn this. Also, the attention paid to the task parts differed between students and experts. Students focused more on the answer choices and less on the text needed. This finding is in accordance with previous research which showed significant inadequacies in students' reading comprehension with regards to their task-solving results (Imam et al., 2014; Akbasli et al., 2016; Tóthová et al., 2021; Tóthová and Rusek, 2021a). It further stresses the aforementioned need to understand students' performance on PISA (and PISA like) tasks better as they have immense impact on education systems, despite the reasons not seeming to be very clear. The most problematic part seen by students was working with tables and maps. This is surprising when visualizations play a crucial role in science as well as in science education (Gilbert, 2005). This phenomenon was not connected only to the general scientific task, but also to the field-specific periodic table. Students spent significantly more time on its legend which described the groups' names, despite it not being relevant for solving the task.

Another reason for the different solving process may be the fact that a student's brain cannot solve problems in this manner without memorizing it (Hartman and Nelson, 2015). Research (Tóthová and Rusek, 2021b) showed that supporting problem-solving strategies in several problematic areas, especially careful reading, identifying the main problem and supporting information led to better results and problem-solving skills' development.

Students in this study mentioned memorized information to be a determinant of successful chemistry task solution. This is in contrast with the fact they could find the information in the tasks. This was later confirmed in the interviews, where reading and working with information was named as the most difficult. The results therefore showed a need to present chemistry tasks in more variable ways to pre-service teacher students and explain their reasoning beside testing field-specific, separated, memorized information.

The results of this study are affected by two major factors which limit the extent to which the findings can be generalized. First, it is the small sample. On the one hand, it provides information only about a small group of participants, on the other, it enables the use of a vast palette of interconnected methods which enable a thorough description of the studied phenomenon. Based on these data, it is later possible to focus on a smaller, more concrete aspect on a larger research sample. Second, it is the sample selection. Though the students were pre-selected based on their pre-test result, the performance especially of partially-successful and unsuccessful students was quite similar which did not result in as many findings as expected. Again, with a larger sample, more differences could be found. Also, as the experts sample was convenient, further research could include also experts from science disciplines.

Conclusion

The comparison of students and experts' general science and chemistry oriented complex problem tasks showed several trends in the participants' problem-solving processes. Combining tasks, eye-tracking, cued retrospective think-aloud and interviews, though time consuming, brought several important findings which deepen contemporary understanding of the problem-solving process. Despite the results not being generalized due to a small sample, they have the potential to inform the (science) education community in its endeavor to more effective instruction.

The pre-service teacher students considered chemistry tasks more difficult than science tasks, which was reflected in their results. On the other hand, the experts considered science tasks more complex and more difficult. The reasons behind the differences in the groups' performance revealed possible areas the students need to improve but also raised more questions to be answered in future research.

The experts spent more time reading the task context in the PISA task requiring general scientific knowledge. However, their time-fixation duration was shorter in most parts of the chemistry-related task that proved their expertise. On the contrary, students' longer time spent on the unnecessary information was one of the indicators for their lower success. Their task-solving processes revealed their lower ability to use information provided in the text, which was identified through the lack of attention paid to the periodic table and confirmed in their spoken description of the problem-solving process (think-aloud and interviews). In the interviews, even successful students tended to mention the importance of memorized information, mainly in chemistry tasks (the field of their study), and, that provided information confused them *"you simply have to know that."*

The differences between the students and experts were also shown in the strategies they used. Both groups used mainly expansive strategies. However, students used limiting strategies 22% of the time, unlike experts who did not apply limiting strategies at all. The expertness consisted in the participants' variability of applied expansive strategies even in cases where the originally chosen strategy did not work. This is another issue teacher training needs to focus on.

Limiting strategies were connected to the problems the solvers faced. Logical reasoning, knowledge, and understanding an assignment were proved to be crucial. When reading a task, differences between reading behavior in students and experts appeared to be a possible reason for incorrect solutions. Therefore field-specific reading needs to be focused on pre-service teacher training.

The results also showed a need to support pre-service teachers' ability to identify the main problem and supporting information. Moreover, the results showed a need to present chemistry tasks in more variable ways to PCTs and explain their reasoning other than testing field-specific, separated, memorized information. As the pool of identified novice vs. expert differences is already quite full, future research should focus on specific means for effective procedure transfer to students. As it includes many hidden processes, the combination of methods used in this study (eye-tracking, think-aloud, and interviews) are methods which will surely find their use.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

MR was responsible for the task selection and evaluation. MT was responsible for the eye-tracking data gathering and analysis. The results' discussion, their implications as well as article writing was done in cooperation. Both authors contributed to the publication equally. Together, they shaped the study design.

Funding

This study was supported by the Charles University Grant Agency grant no. 492122. Analysis of the pre-service chemistry teachers' process of solving tasks with problematic elements and the project Cooperation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Akbasli, S., Sahin, M., and Yaykiran, Z. (2016). The Effect of Reading Comprehension on the Performance in Science and Mathematics. *J. Educ. Pract.* 7, 108–121.

Bakken, J., and Andersson-Bakken, E. (2021). The textbook task as a genre. *J. Curric. Stud.* 53, 729–748. doi: 10.1080/00220272.2021.1929499

Barba, R. H., and Rubba, P. A. (1992). A comparison of preservice and in-service earth and space science teachers' general mental abilities, content knowledge, and problem-solving skills. *J. Res. Sci. Teach.* 29, 1021–1035. doi: 10.1002/tea.3660291003

Barham, A. I. (2020). Investigating the Development of Pre-Service Teachers' Problem-Solving Strategies via Problem-Solving Mathematics Classes. *Eur. J. Educ. Res.* 9, 129–141.

Bellanca, J. A. (2010). *21st century skills: Rethinking how students learn.* Bloomington: Solution Tree Press.

Ben-Zvi, N., and Genut, S. (1998). Uses and limitations of scientific models: The Periodic Table as an inductive tool. *Int. J. Sci. Educ.* 20, 351–360. doi: 10.1080/0950069980200307

Caleon, I., and Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *Int. J. Sci. Educ.* 32, 939–961. doi: 10.1080/09500690902890130

Cheung, D. (2009). Using think-aloud protocols to investigate secondary school chemistry teachers' misconceptions about chemical equilibrium. *Chem. Educ. Res. Pract.* 10, 97–108. doi: 10.1039/B908247F

De Wever, B., Hämäläinen, R., Nissinen, K., Mannonen, J., and Van Nieuwenhove, L. (2021). Teachers' problem-solving skills in technology-rich environments: A call for workplace learning and opportunities to develop professionally, Teachers' problem-solving skills in technology-rich environments: A call for workplace learning and opportunities to develop professionally*. *Stud. Cont. Educ.* 1–27. doi: 10.1080/0158037X.2021.2003769

Gilbert, J. K. (2005). "Visualization: A metacognitive skill in science and science education," in *Visualization in science education. Models and Modeling in Science Education*, ed. J. K. Gilbert (Dordrecht: Springer), 9–27.

Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *Eur. Educ. Res. J.* 15, 598–616. doi: 10.1177/1474904116658299

Harsh, J. A., Campillo, M., Murray, C., Myers, C., Nguyen, J., and Maltese, A. V. (2019). "Seeing" Data Like an Expert: An Eye-Tracking Study Using Graphical Data Representations. *CBE Life Sci. Educ.* 18:ar32. doi: 10.1187/cbe.18-06-0102

Hartman, J. R., and Nelson, E. A. (2015). "Do we need to memorize that?" or cognitive science for chemists. *Found. Chem.* 17, 263–274. doi: 10.1007/s10698-015-9226-z

Holec, J., and Rusek, M. (2016). *Metodické komentáøe a úlohy ke Standardùm pro základní vzdìlávání - chemie*. Praha: NÚV.

Imam, O. A., Mastura, M. A., Jamil, H., and Ismail, Z. (2014). Reading comprehension skills and performance in science among high school students in the Philippines. *Asia Pac. J. Educ. Educ.* 29, 81–94.

Jian, Y.-C. (2018). Reading instructions influence cognitive processes of illustrated text reading not subject perception: An eye-tracking study. *Front. Psychol.* 9:2263. doi: 10.3389/fpsyg.2018.02263

Klein, P., Viiri, J., Mozaffari, S., Dengel, A., and Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots?. *Phys. Rev. Phys. Educ. Res.* 14:010116. doi: 10.1103/PhysRevPhysEducRes.14.010116

Kreiner, S., and Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika* 79, 210–231. doi: 10.1007/s11336-013-9347-z

Krulik, S., and Rudnick, J. A. (1982). Teaching problem solving to preservice teachers. *Arithmetic Teach.* 29, 42–45. doi: 10.5951/AT.29.6.0042

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., et al. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* 10, 90–115. doi: 10.1016/j.edurev.2013.10.001

Lee, K. W. L., Tan, L. L., Goh, N. K., Chia, L. S., and Chin, C. (2000). Science teachers and problem solving in elementary schools in Singapore. *Res. Sci. Technol. Educ.* 18, 113–126. doi: 10.1080/713694953

Liampa, V., Malandrakis, G. N., Papadopoulou, P., and Pnevmatikos, D. (2019). Development and Evaluation of a Three-Tier Diagnostic Test to Assess Undergraduate Primary Teachers' Understanding of Ecological Footprint. *Res. Sci. Educ.* 49, 711–736. doi: 10.1007/s11165-017-9643-1

Lindner, M. A., Eitel, A., Strobel, B., and Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learn. Instruct.* 47, 91–102. doi: 10.1016/j.learninstruc.2016.10.007

Mandíková, D., and Houfková, J. (2012). *Úlohy pro rozvoj pøírodovìdné gramotnosti*. Praha: ÈŠI.

Medková, I. (2013). *Dovednosti žákù ve výuce fyziky na základní škole [The Pupil's Skills in the Physics Education at the Lower-secondary School]*. Brno: Masarykova univerzita.

Milbourne, J., and Wiebe, E. (2018). The Role of Content Knowledge in Ill-Structured Problem Solving for High School Physics Students. *Res. Sci. Educ.* 48, 165–179. doi: 10.1007/s11165-016-9564-4

OECD (2016). *Pisa 2015 Results (Volume 1): Excellence and Equity in Education*. Paris: OECD Publishing.

OECD (2018). *The Future of education and skills*. Paris: OECD Publishing.

OECD (2019). *PISA 2018 Results (Volume I)*. Paris: OECD Publishing.

Ogilvie, C. A. (2009). Changes in students' problem-solving strategies in a course that includes context-rich, multifaceted problems. *Phys. Rev. Special Top. Phys. Educ. Res.* 5:020102. doi: 10.1103/PhysRevSTPER.5.020102

Osterhage, J. L. (2021). Persistent miscalibration for low and high achievers despite practice test feedback in an introductory biology course. *J. Microbiol. Biol. Educ.* 22, e00139–21. doi: 10.1128/jmbe.00139-21

Phelps, A. J. (1996). Teaching to enhance problem solving: It's more than the numbers. *J. Chem. Educ.* 73:301. doi: 10.1021/ed073p301

Potužníková, E. Lokajíèková, V., and Janík, T. (2014). Mezinárodní srovnávací výzkumy školního vzdìlávání v Èeské republice: Zjištìní a výzvy. *Pedagogická Orientace* 24, 185–221. doi: 10.5817/PedOr2014-2-185

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* 6, 135–147.

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2020). Decoding Case Comparisons in Organic Chemistry: Eye-Tracking Students'. Visual Behavior. *J. Chem. Educ.* 97, 3530–3539. doi: 10.1021/acs.jchemed.0c00418

Rusek, M., Koreneková, K., and Tóthová, M. (2019). "How Much Do We Know about the Way Students Solve Problem-tasks," in *Project-based Education and Other Activating Strategies in Science Education XVI*, eds M. Rusek and K. Vojíø (Prague: Charles University, Faculty of Education).

Rusek, M., and Tóthová, M. (2021). "Did students reach the periodic table related curricular objectives after leaving from lower-secondary education?," in *Project-Based Education and Other Activating Strategies in Science Education XVIII*, eds M. Rusek, M. Tóthová, and K. Vojíø (Prague: Charles University, Faculty of Education).

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., and Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *J. Educ. Psychol.* 106, 974–989. doi: 10.1037/a0037054

Son, J. W., and Kim, O. K. (2015). Teachers' selection and enactment of mathematical problems from textbooks. *Math. Educ. Res. J.* 27, 491–518. doi: 10.1007/s13394-015-0148-9

Štech, S. (2018). PISA–nástroj vzdìlávací politiky nebo výzkumná metoda? [PISA – An Instrument of Educational Policy or A research Method?]. *Orbis Sch.* 5, 123–133. doi: 10.14712/23363177.2018.81

Stenlund, T., Eklöf, H., and Lyrén, P. E. (2017). Group differences in test-taking behaviour: An example from a high-stakes testing program. *Assess. Educ.* 24, 4–20. doi: 10.1080/0969594X.2016.1142935

Straková, J. (2016). *Mezinárodní výzkumy výsledkù vzdìlávání: Metodologie, pøínosy, rizika a pøíležitosti [International research on educational outcomes: Methodology, benefits, risks and opportunities]*. Prague: Univerzita Karlova.

Tai, R. H., Loehr, J. F., and Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *Int. J. Res. Method Educ.* 29, 185–208. doi: 10.1080/17437270600891614

Talsma, K., Schüz, B., and Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy? self-fulfilling prophecy. *Learn. Individ. Differ.* 69, 182–195. doi: 10.1016/j.lindif.2018.11.002

Topczewski, J. J., Topczewski, A. M., Tang, H., Kendhammer, L. K., and Pienta, N. J. (2017). NMR spectra through the eyes of a student: Eye tracking applied to NMR items. *J. Chem. Educ.* 94, 29–37. doi: 10.1021/acs.jchemed.6b00528

Tóthová, M., and Rusek, M. (2021b). Developing Students' Problem-solving Skills Using Learning Tasks: An Action Research Project in Secondary School. *Acta Chim. Slov.* 68, 1016–1026. doi: 10.17344/acsi.2021.7082

Tóthová, M., and Rusek, M. (2021c). The Use of Eye-tracking in Science Textbook Analysis: A Literature Review. *Sci. Educ.* 12, 63–74. doi: 10.14712/18047106.1932

Tóthová, M., and Rusek, M. (2021a). "An analysis of pre-service chemistry teachers' progress when solving multicomponent tasks," in *Project-Based Education and Other Activating Strategies in Science Education XVIII*, eds M. Rusek, M. Tóthová, and K. Vojíø (Prague: Charles University, Faculty of Education).

Tóthová, M., Rusek, M., and Chytrý, V. (2021). Students' Procedure When Solving Problem Tasks Based on the Periodic Table: An Eye-Tracking Study. *J. Chem. Educ.* 98, 1831–1840. doi: 10.1021/acs.jchemed.1c00167

Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., and Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Comput. Educ.* 58, 375–385. doi: 10.1016/j.compedu.2011.07.012

van der Graaf, J., Segers, E., and de Jong, T. (2020). Fostering integration of informational texts and virtual labs during inquiry-based learning. *Contemp. Educ. Psychol.* 62:101890. doi: 10.1016/j.cedpsych.2020.101890

Van Gog, T., Paas, F., Van Merriënboer, J. J., and Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *J. Exp. Psychol. Appl.* 11:237. doi: 10.1037/1076-898X.11.4.237

Van Someren, M., Barnard, Y., and Sandberg, J. (1994). *The think aloud method: A practical approach to modelling cognitive*. London: London Academic Press.

Vega Gil, L., Beltrán, J. C. H., and Redondo, E. G. (2016). PISA as a political tool in Spain: Assessment instrument, academic discourse and political reform. *Eur. Educ.* 48, 89–103. doi: 10.1080/10564934.2016.1185685

Vojíø, K., and Rusek, M. (2022). Opportunities for learning: Analysis of Czech lower-secondary chemistry textbook tasks. *Acta Chim. Slov.* 69, 359–370. doi: 10.17344/acsi.2021.7245

Wu, C. J., and Liu, C. Y. (2021). Eye-movement study of high-and low-prior-knowledge students' scientific argumentations with multiple representations. *Phys. Rev. Phys. Educ. Res.* 17:010125. doi: 10.1103/PhysRevPhysEducRes.17.010125

Zhao, Y. (2020). Two decades of havoc: A synthesis of criticism against PISA. *J. Educ. Change* 21, 245–266. doi: 10.1007/s10833-019-09367-x

TYPE Correction
PUBLISHED 27 June 2023
DOI 10.3389/feduc.2023.1241619

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Martin Rusek
martin.rusek@pedf.cuni.cz

RECEIVED 16 June 2023
ACCEPTED 19 June 2023
PUBLISHED 27 June 2023

CITATION
Tóthová M and Rusek M (2023) Corrigendum: "Do you just have to know that?" Novice and experts' procedure when solving science problem tasks. *Front. Educ.* 8:1241619.
doi: 10.3389/feduc.2023.1241619

Corrigendum: "Do you just have to know that?" Novice and experts' procedure when solving science problem tasks

Martina Tóthová and Martin Rusek*

Department of Chemistry and Chemistry Education, Faculty of Education, Charles University, Prague, Czechia

KEYWORDS

problem-solving skills, chemistry education, science education, eye-tracking, pre-service teachers

A corrigendum on
"Do you just have to know that?" Novice and experts' procedure when solving science problem tasks

by Tóthová, M., and Rusek, M. (2022). *Front. Educ.* 7:1051098. doi: 10.3389/feduc.2022.1051098

In the published article, there was an error in Figure 1 as published. There was a shift of images. The picture from the original Figure 8 should be there (resp. Figure 1 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 2 as published. There was a shift of images. The picture from the original Figure 1 should be there (resp. Figure 2 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 3 as published. There was a shift of images. The picture from the original Figure 2 should be there (resp. Figure 3 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 4 as published. There was a shift of images. The picture from the original Figure 3 should be there (resp. Figure 4 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 5 as published. There was a shift of images The picture from the original Figure 4 should be there (resp. Figure 5 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 6 as published. There was a shift of images The picture from the original Figure 5 should be there (resp. Figure 6 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 7 as published. There was a shift of images The picture from the original Figure 6 should be there (resp. Figure 7 in the attached files). The corrected figure and its caption appears below.

In the published article, there was an error in Figure 8 as published. There was a shift of images The picture from the original Figure 7 should be there (resp. Figure 8 in the attached files). The corrected figure and its caption appears below.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

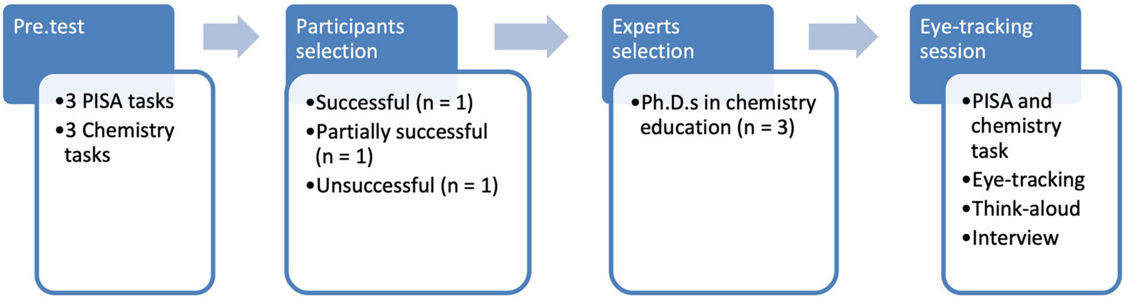


FIGURE 1
The study design.

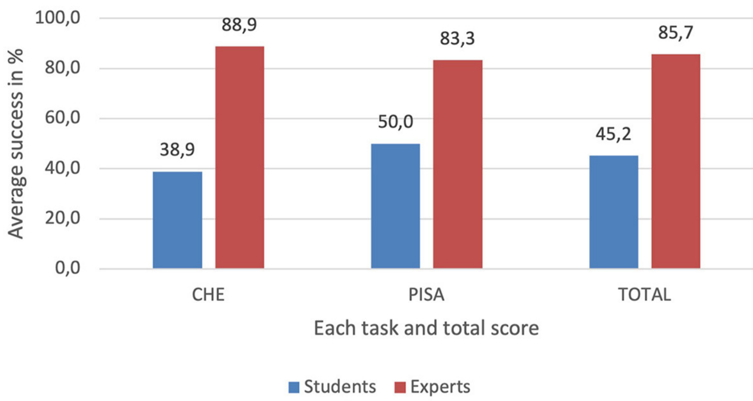


FIGURE 2
Participants' performance in the program of international student assessment (PISA) and chemistry (CHE) problem tasks.

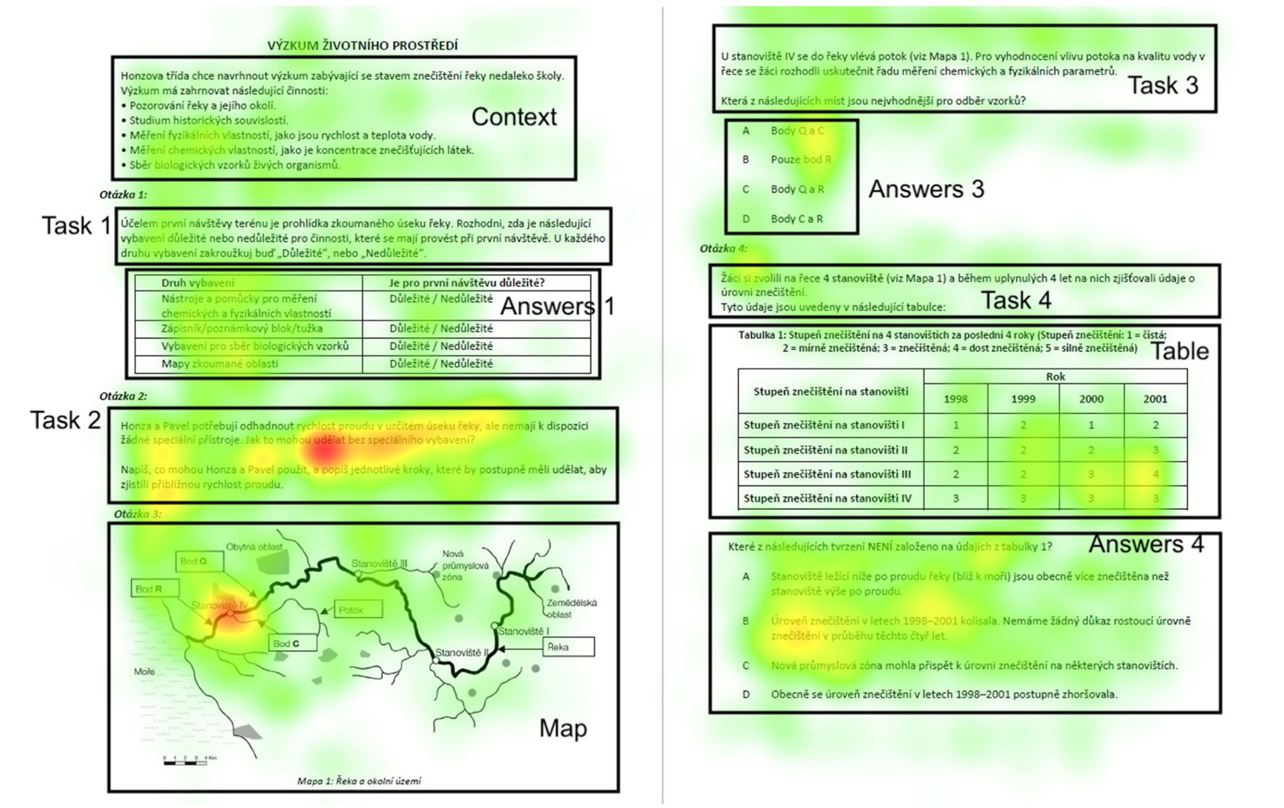


FIGURE 3
Experts' focus in the program of international student assessment (PISA) task.

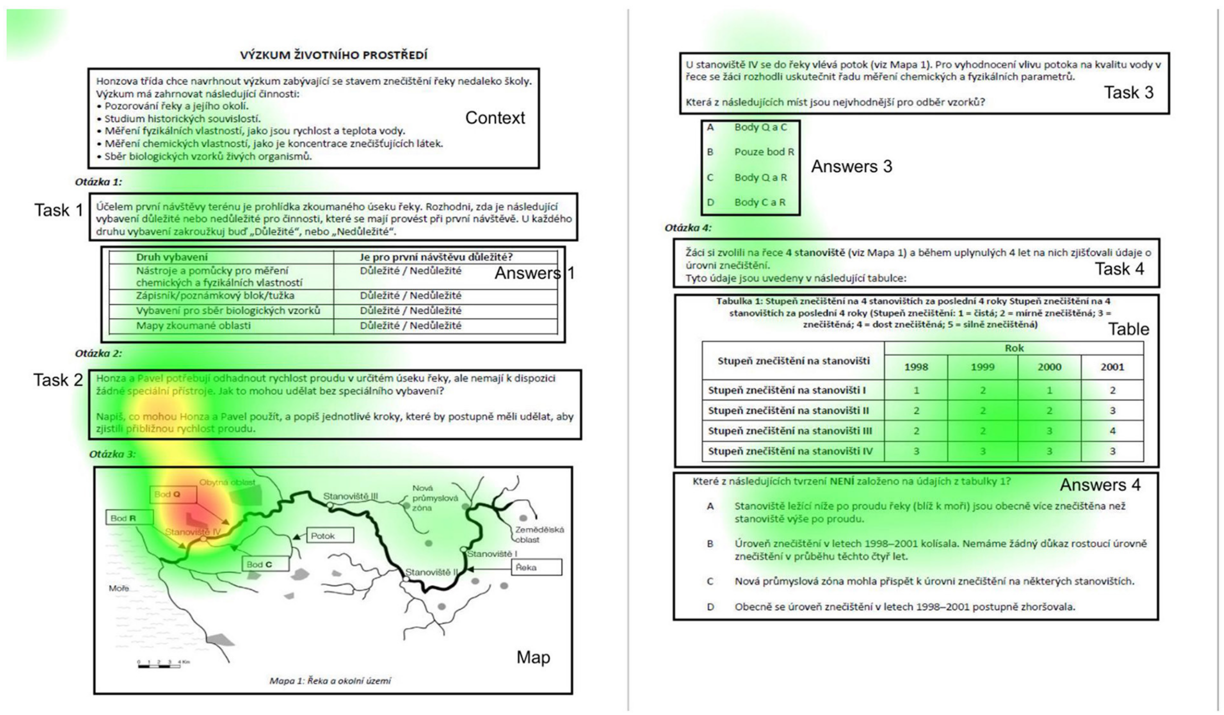


FIGURE 4
Students' focus in the program of international student assessment (PISA) task.

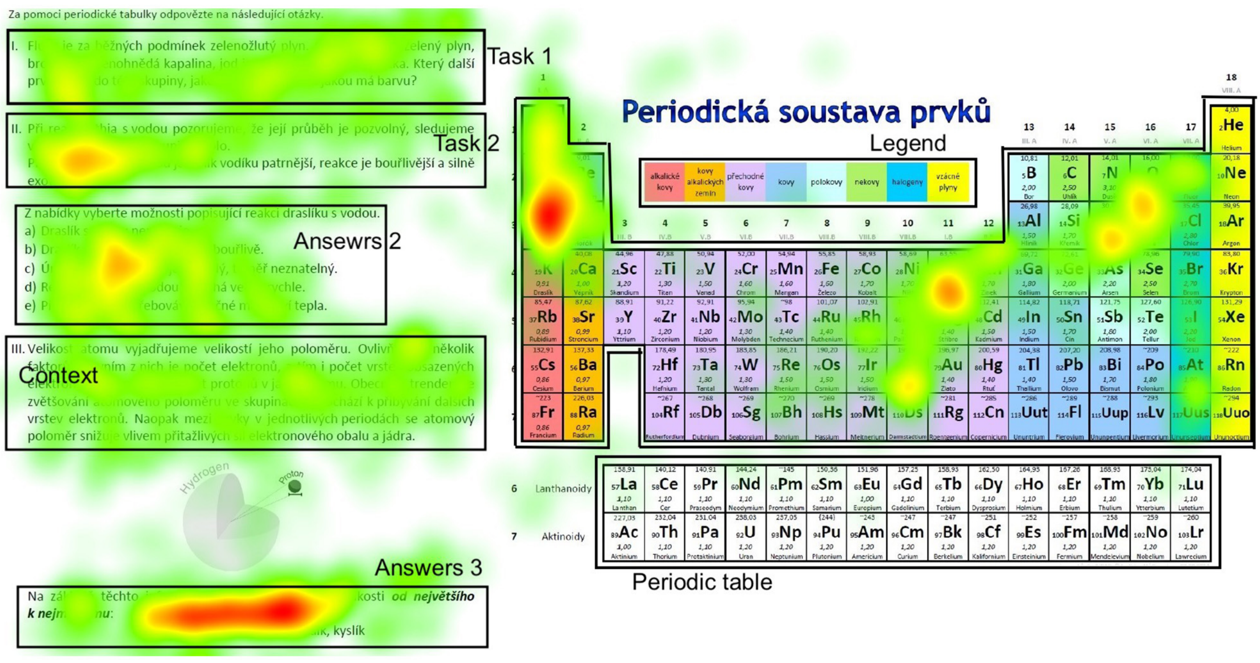


FIGURE 5
Experts' focus in the chemistry task.

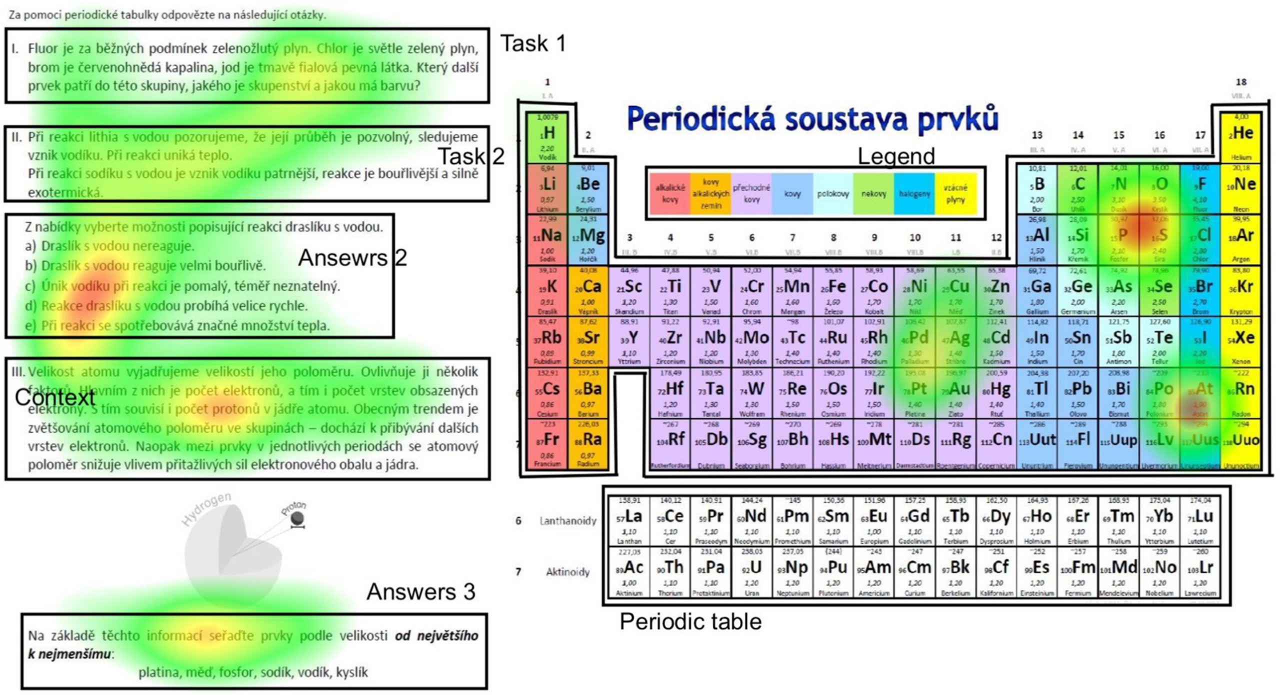


FIGURE 6
Students' focus in the chemistry task.

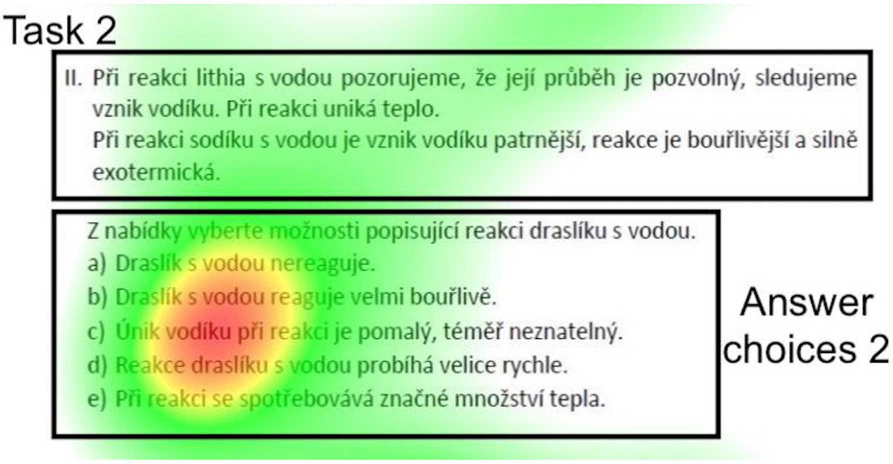


FIGURE 7
Students' attention to the answer choices.

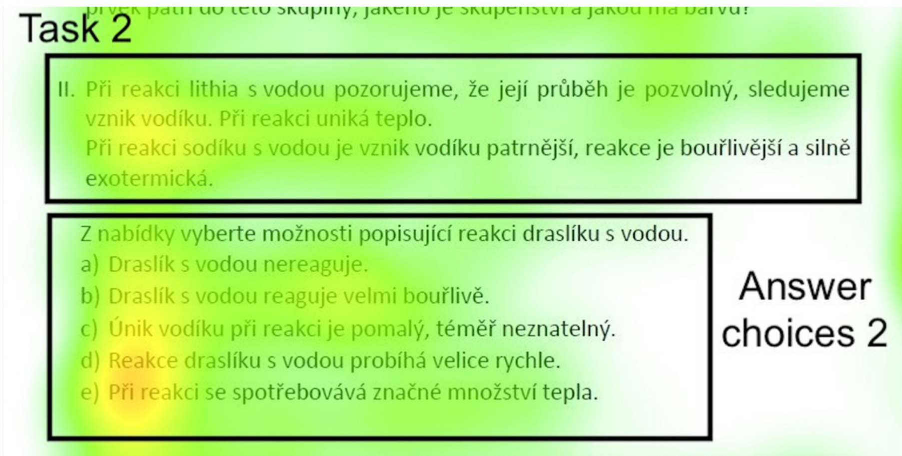


FIGURE 8
Expert's attention to the task itself and answer choices.

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 09 December 2022
DOI 10.3389/feduc.2022.1055280

Check for updates

OPEN ACCESS

EDITED BY
Pascal Klein,
University of Göttingen, Germany

REVIEWED BY
Sandra Gudino Paredes,
Monterrey Institute of Technology and Higher Education, Mexico
Dihua Xue,
University of Minnesota Rochester, United States
Solaire Finkenstaedt-Quinn,
University of Michigan, United States

*CORRESPONDENCE
Nicole Graulich
Nicole.Graulich@didaktik.chemie.uni-giessen.de

SPECIALTY SECTION
This article was submitted to STEM Education, a section of the journal Frontiers in Education

RECEIVED 27 September 2022
ACCEPTED 21 November 2022
PUBLISHED 09 December 2022

CITATION
Braun I, Langner A and Graulich N (2022) Let's draw molecules: Students' sequential drawing processes of resonance structures in organic chemistry.
Front. Educ. 7:1055280.
doi: 10.3389/feduc.2022.1055280

Let's draw molecules: Students' sequential drawing processes of resonance structures in organic chemistry

Irina Braun, Axel Langner and Nicole Graulich*

Department of Biology and Chemistry, Institute of Chemistry Education, Justus Liebig University Giessen, Giessen, Germany

Drawing is a fundamental skill in science, technology, engineering, and mathematics (STEM) disciplines to express one's reasoning and externalize mental models in problem-solving. Although research has highlighted the effectiveness of drawing as a learning strategy and the importance of drawing accuracy for learning success, little is known about learners' actual drawing process. However, especially in organic chemistry, the investigation of drawing processes is of great importance as generating different representations, such as structural formulas, is inherent to problem-solving in this visual-laden discipline. Resonance structures, for example, are often used to estimate reactive sites in a molecule and to propose reaction pathways. However, this type of representation places a high cognitive demand on learners, which, besides conceptual difficulties, leads to drawing difficulties. To support learners in drawing and using resonance structures in problem-solving, it is necessary to characterize how they generate their drawings. To this end, a qualitative, exploratory study has been conducted to investigate undergraduate students' (N = 20) drawing processes of resonance structures while solving an organic case comparison task. Using eye-tracking, the characteristics regarding the construction of productive and unproductive drawings became visible. Results indicate that unproductive drawings often stem from integrating and connecting unrelated information during the drawing process. Further, the results show that the productivity of a drawing depends on learners' flexibility in information selection. Implications for supporting learners' drawing process and using eye-tracking for characterizing drawing processes in other STEM disciplines are discussed.

KEYWORDS

eye-tracking, organic chemistry, drawing, resonance, chemistry education research (CER), undergraduate, molecular structures

10.3389/feduc.2022.1055280

Introduction

In science, technology, engineering, and mathematics (STEM) disciplines, scientists and learners rely heavily on external representations to make sense of scientific concepts and phenomena (Fiorella and Zhang, 2018; Ainsworth and Scheiter, 2021). As representations constitute a fundamental means for the construction and transmission of knowledge, students need to be proficient in analyzing and using given representations (Nitz et al., 2014) and be fluent in generating them (Ainsworth et al., 2011). In fact, by expressing one's reasoning and externalizing mental models, drawings can help learning new concepts and support problem-solving (Cox, 1999; Ainsworth et al., 2011; Quillin and Thomas, 2015; Cooper et al., 2017; Wu and Rau, 2019). In recent years, much research has been conducted on drawing as a learning activity across different STEM contexts such as chemistry (Hellenbrand et al., 2019), physics (Maries and Singh, 2018), biology (Schmeck et al., 2014), and geography (Gobert and Clement, 1999) at school and college levels. Predominantly, it has been shown that prompting learners to visually depict content presented in text- or animation-based instructional materials is an effective learning strategy as students who draw not only build higher quality explanations and develop a more coherent mental model of a studied phenomenon, but also perform better in subsequent tasks and tests (Bobek and Tversky, 2016; Fiorella and Zhang, 2018; Cromley et al., 2019). Thus, drawing enhances learning outcomes related to retention, comprehension, and knowledge transfer (Van Meter and Garner, 2005; Leopold and Leutner, 2012; Schmeck et al., 2014; Fiorella and Zhang, 2018; Fiorella et al., 2020). However, drawing *per se* does not automatically enhance learning. By referring to the prognostic drawing principle, Schwamborn et al. (2010) point out that the quality of learners' drawings is predictive of their learning outcomes, i.e., the more accurately and correctly learners draw, the better their performance. This finding has been replicated in various studies (Mason et al., 2013; Schmeck et al., 2014; Rellensmann et al., 2016; Fiorella and Zhang, 2018; Hellenbrand et al., 2019; Schmidgall et al., 2020; Stieff and DeSutter, 2020).

Despite focusing on the quality of final drawing products and their relation to learning outcomes, the actual drawing process leading to these products has not received much attention in research so far (Lobato et al., 2014). However, drawing constitutes a core scientific practice itself and, consequently, necessitates a profound understanding of how a drawing is sequentially generated and which factors influence its quality. In organic chemistry, for instance, drawing is fundamental to scientific thinking and model-based reasoning as diagrams and structural formulas have a high explanatory power and make imperceptible entities and processes visible (Goodwin, 2008; Cooper et al., 2017; Graulich and Bhattacharyya, 2017). Whether in the laboratory or the lecture, chemists sketch and manipulate molecular structures to explain findings and communicate chemical content (Kozma et al., 2000). Besides this displaying function, drawings of molecular structures are crucial problem-solving tools. Since much chemical information (e.g., connectivity, polarity) is embedded within molecular structures, drawing enables the expression of assumptions or predictions about the properties of molecules or possible reaction processes (Cartrette and Bodner, 2009; Cooper et al., 2017). Constructing resonance structures of organic molecules, for instance, serves as a mean to estimate reactive sites in a molecule by representing multiple variants of the electronic delocalization in a molecule which cannot be adequately represented by a single structure. By considering the hypothetical electronic distribution in a molecular structure (i.e., the contribution of each resonance structure to the dynamic electron density distribution of a molecule), resonance structures enable the prediction of reaction pathways (Richardson, 1986). Therefore, constructing appropriate resonance structures constitutes the first critical step in a chain of steps of inferences to derive structural properties and chemical reactivities (Cooper et al., 2012). Thus, not succeeding in this step hinders students from using the resonance structures adequately in subsequent problem-solving (Strickland et al., 2010; Carle and Flynn, 2020). To support students in the adequate use of the resonance concept and the respective drawing process of resonance structures, it is crucial to understand how students are sequentially generating their drawings and which drawing behavior characterizes the generation of productive, thus valid and significant resonance structures, and unproductive drawings, i.e., wrong or insignificant resonance structures.

A more profound, process-oriented characterization of drawing processes can be achieved by using eye-tracking. Without interfering with the construction process, recording eye movements quantitatively captures students' visual processing of stimuli (e.g., by providing insights into learners' attention distribution on structural features or their search behavior in terms of gaze patterns). This can help to draw conclusions about learners' underlying cognitive processes and different strategic approaches when constructing representations such as molecular structures (Just and Carpenter, 1980; Cullipher et al., 2018). Therefore, using eye-tracking in the context of construction processes in STEM disciplines allows a deeper insight into students' drawing processes and offers a new perspective on obstacles students encounter when generating representations such as chemical structures.

Prior research on students' difficulties in the construction of organic molecular structures

Becoming proficient in using symbolic language in organic chemistry, such as drawing mechanisms or using the electron-pushing formalism, has been constantly shown to be difficult for students at various levels (Bodner and Domin, 2000; Graulich, 2015; Flynn and Featherstone, 2017; Dood and Watts, 2022). Several studies report that students depict mechanisms solely based on rote memorization, do not ascribe meaning to the electron-pushing formalism, place electron arrows as decoration instead, and exhibit difficulties in the construction of structural formulas (Bhattacharyya and Bodner, 2005; Cooper et al., 2010; Grove et al., 2012). Lewis structures are typically the first type of representations students encounter in organic chemistry to represent molecular structures as a variety of physical and chemical properties of molecular compounds can be inferred by constructing and inspecting Lewis structures (e.g., physical state, geometry, solubility) (Cooper et al., 2012; Tiettmeyer et al., 2017). In fact, drawing Lewis structures implies a high cognitive load on students, as students must consider various concepts (e.g., expanded octets, geometry) and sets of rules while coping with many exceptions to these very rules at the same time (Cooper et al., 2010; Kaufmann et al., 2017; Tiettmeyer et al., 2017; Karonen et al., 2021). Cooper et al. (2010) intensively investigated students' ability to construct and manipulate Lewis structures across different learning levels. As a main finding, they demonstrated that students' and even faculty members' competence in constructing valid Lewis structures is deficient. Most students struggled with creating valid Lewis structures involving two or more carbon atoms (Cooper et al., 2010). Moreover, it became apparent that the success rate depended on how the formulas were presented to students. While students struggled with drawing a Lewis structure of methanol in the form of CH_4O, more students could produce a correct structure if the functional group was depicted explicitly, i.e., as CH_3OH (Cooper et al., 2010). Other common errors that students exhibit when drawing Lewis structures encompass the inability to determine the correct number of bonds, the arbitrary assignment of (formal) charges in ions, or the overreliance on rules such as the octet rule (Cooper et al., 2010; Kaufmann et al., 2017; Karonen et al., 2021). Concerning the latter, students were either likely to apply the octet rule when it was not possible to show atoms with full octets or tended to violate the octet rule, e.g., by exceeding the octet or depicting atoms different from carbon (e.g., nitrogen or oxygen atoms) as electron-deficient (Cooper et al., 2010). Some students in Cooper et al.'s (2010) study even invented their own rules and invalid strategies (e.g., reaching the highest symmetry). All this suggests that students tend to rely on memorized, salient cues while drawing, use rules mechanically, or approach drawing tasks by unsystematic trial-and-error strategies (Ahmad and Omar, 1992; Cooper et al., 2010; Kaufmann et al., 2017; Sandi-Urena et al., 2020).

These problems reported for the generation of single Lewis structures also apply to the construction of resonance structures. Resonance structures provide a more accurate way of describing molecules that single Lewis structures cannot accurately display due to the delocalization of π-electron pairs over several atoms. As such, molecules can be best described by drawing multiple structural formulas (e.g., Lewis structures or skeletal formulas) of the same molecule with a varying electron distribution. The combination of these drawings, depending on their contribution to the overall stability of the molecule, ultimately forms the resonance hybrid of the given molecule. Consequently, the properties and reactivities (e.g., charge density and product distribution) of molecules that exhibit resonance lie between the different canonical structures. Although students are expected to have a thorough understanding of this core chemical concept after introductory courses in organic chemistry and should be able to use it fluently to depict the electronic structure of compounds, research in chemistry education has demonstrated that the resonance concept puts a high cognitive load on students, even at the university level (Duis, 2011; Brandfonbrener et al., 2021). This leads to various misconceptions, such as considering resonance structures as equilibrium or electron reservoirs (Taber, 2002; Kim et al., 2019; Xue and Stains, 2020). As the application of this concept requires the integration of different concepts and prior knowledge (e.g., electronegativity, hybridization, electron-pushing formalism), a fragmented conceptualization of these relationships may hinder subsequent problem-solving (Betancourt-Pérez et al., 2010). Besides studies focusing on students' conceptual understanding of resonance, little research has been conducted on how students draw resonance structures. For instance, Betancourt-Pérez et al. (2010), used different tasks to investigate students' competence in the construction of resonance structures across different learning levels in organic chemistry by prompting their participants to (a) draw curved arrows to show the electron movement in resonance structures, (b) draw alternative structures for a given ion or molecule, (c) identify the most stable resonance structure, and (d) draw the resonance hybrid. Their results show that students perform poorly in drawing resonance structures and exhibit different errors in both the first and second semester. The most prevalent errors encompass, for example, the violation of the octet rule, e.g., by moving π-bonds toward atoms with a full octet, irrespective of the atom's hybridization and number of bonds. Moreover, students tend to break σ-bonds between carbon and hydrogen atoms, put charges on atoms that are not charged or construct resonance structures with a different delocalized system, thus different connectivity, compared to the initial structure (Betancourt-Pérez et al., 2010). The authors concluded that students, especially at the beginning of their studies, do not pay much attention to details when drawing resonance structures (Betancourt-Pérez et al., 2010). In a recent study, Petterson et al. (2020) demonstrated that students struggled to identify the correct place to start the movement of electrons when deriving one resonance structure from another in the context of acid-base reaction mechanisms. Although these results may be explained as knowledge gaps related to the resonance concept, it remains unclear what actually characterizes students' drawing process, i.e., what structural features learners perceptually pay attention to when translating one resonance structure into another, eventually leading to invalid structures and therefore causing wrong or erroneous inferences. Thus, students' ability to decode and manipulate molecular structures requires a more profound analysis to determine the sources of students' success or difficulties while drawing.

Theoretical framework

Representational competence

Learning chemistry includes learning how to effectively use representations such as chemical symbols, diagrams, or ball-and-stick-models to make "sense of the invisible and untouchable" (Kozma and Russell, 1997, p. 949) and, consequently, connect the molecular level to the corresponding macroscopic phenomenon (Johnstone, 1991). Often, multiple representations are combined to facilitate learning and problem solving, as they can either complement each other by offering different perspectives of the given phenomenon, constraint the interpretation of the provided material, or help in constructing a more profound understanding (e.g., by facilitating abstraction) (Ainsworth, 2006). Depending on the coding system of the representations in the working memory channel (i.e., symbolic or analogous), one can differentiate between multiple heterogeneous (i.e., a combination of symbolic and analogous representations) and homogeneous (i.e., either exclusively symbolic or exclusively analogous representations) representations (Ott et al., 2018; Malone et al., 2020). While a heterogeneous representational system in chemistry could be, for example, the combination of a drawn Lewis structure and a ball-and-stick model, a homogeneous system could be the combination of a text and a reaction equation or, more specifically, the depiction of multiple resonance structures. Irrespective of the combination, to profit from (multiple) representations and their synergies for knowledge acquisition and problem-solving, learners need to develop representational literacy (Lesh et al., 1987; Gilbert, 2005; Cooper et al., 2017). This means not only becoming proficient in the use of each representation, but also being able to interrelate corresponding information within and across these different representations and, thus, construct a coherent mental representation of the phenomenon (i.e., local and global coherence formation) (Seufert, 2003; Seufert and Brünken, 2006). Concerning the construction and use of resonance structures, for instance, the relationship between different structures, i.e., how one structure can be transformed into another, and their specific contribution to the overall electronic distribution in a molecule, is crucial.

Based on the comparison of expert and novice representational practices, Kozma and Russell (1997, 2005) defined representational competence by deriving a set of cognitive skills and practices that enable a person to successfully use representations to reason about, express ideas, and create meaning for scientific phenomena. In particular, representational competence encompasses the abilities to (1) use representations to describe observable chemical phenomena and their underlying molecular entities and processes, (2) select or generate a representation and explain its appropriateness for a particular purpose, (3) identify, describe and interpret features of a particular representation, (4) make connections across different related representations by mapping features of one representation onto those of another, and explain the relationship between them, (5) take the epistemological position that representations are modeling but are distinct from the phenomenon observed, (6) evaluate representations by describing limitations and affordances of different representations, and (7) use representations in social situations to support claims, draw inferences, and make predictions about chemical phenomena (Kozma and Russell, 1997, 2005). These skills follow a developmental trajectory, i.e., they develop in sophistication over time. Different proficiency levels can be achieved in varying contexts with different types of representations. For instance, a learner may master most of the skills listed above regarding a particular representation (e.g., Newman projections) but achieve only a low level of representational competence for other representations (e.g., reaction coordinate diagrams) (Kozma and Russell, 2005).

To transform resonance structures into one another, multiple representational skills play a role. First, this transformational process involves analyzing the given starting structure by decoding structural features which have the capacity to delocalize electrons. This skill necessitates a global, holistic view of the structure since the delocalization of electrons can be spread over multiple parts of a molecule and is not limited to a single structural feature. Following this selection process, one structure is translated into another by delocalizing π-electrons and constantly evaluating the hypothetical, resulting structure regarding plausibility and the overall electronic distribution. This can subsequently serve as a basis for predictions of reactions. It becomes evident that the translation in this specific homogeneous representational system does not only require the careful mapping of structural features but crucially depends on the ability to interpret a given structure and possible sources for resonance.

In this regard, the ability to interpret representations further depends on different factors (Schönborn and Anderson, 2008, 2010). According to Schönborn and Anderson (2008, 2010), these factors consist of (1) the external features of the representation (Mode, M), (2) the use of underlying cognitive processes and skills to make sense of a representation (Reasoning, R), and (3) learners' (prior) knowledge of relevant concepts (Conceptual, C). Moreover, these three main factors are intertwined, resulting in four additional factors influencing students' ability to interpret representations. That is the

R-C factor which encompasses students' ability to employ appropriate conceptual knowledge necessary for interpreting the representation, the R-M factor that involves the deciphering and perception of the visual information embedded in the representation, and the C-M factor which describes the propositional scientific knowledge that is transmitted through the explicit features of a representation. This may concern, for example, the complexity and clarity of the representation. Finally, engaging all factors, the C-R-M factor embodies students' ability to successfully interpret a representation by linking their conceptual knowledge to the representational features when decoding information communicated by the representation (Schönborn and Anderson, 2008, 2010). It becomes evident that the interpretation and subsequent construction of resonance structures require a sound conceptual understanding of resonance and highly depend on students' ability to decipher and reason with relevant graphical features of a structural representation. Therefore, the perceptual mechanisms guiding the decoding of the representations and underlying the visuospatial operations when constructing resonance structures must be considered for a holistic characterization of students' approaches when constructing resonance structures.

Mechanisms of visual selection

In order to make sense of the visual information representations convey while organizing and integrating it with prior knowledge and making it subsequently available for higher-order cognitive processes such as reasoning, the visual input has to be filtered to select relevant stimuli (Mayer, 2005; Anderson, 2013). Research on visual search differentiates between three competing mechanisms of visual selection driving the allocation of attention: a stimulus-driven, a goal-driven, and history-driven selection (Awh et al., 2012; Anderson, 2013; Theeuwes, 2019). The stimulus-driven selection is considered a bottom-up process that depends on factors external to the observer, such as the visual salience of the stimulus (e.g., heteroatoms in molecules) that is responsible for the attraction of attention (Theeuwes, 2019). In contrast to that, the goal-driven selection proceeds in a top–down process. Here, the visual search goals and, therefore, the intentional, deliberate control of an observer (e.g., the active search for a specific feature due to rules or prompts) influences the attention to features of a stimulus (Theeuwes, 2019). As a domain-specific prompt or problem task requires where to look or what to attend to, the degree of sophisticated domain knowledge may influence how attention is directed to a given visual input and how it is perceived. In constructing resonance structures, this could be embodied, for example, by carefully examining the fulfillment of the octet rule. The history-driven selection applies when previous experiences drive attentional selection. Thus, information selected in the past affects the way information is selected in subsequent situations. That may encompass, for instance, the probability that features having been repeatedly attended to in the past are more often selected and identified in a given situation (e.g., considering double bonds in the context of resonance structures as they have been often delocalized in previous exercises) (Theeuwes, 2019).

Research questions and hypotheses

Although research indicates that students encounter numerous difficulties when constructing resonance structures and that the construction of resonance structures crucially depends on students' competence to deal with structural formulas in terms of decoding, selecting, and manipulating these representations (Kozma and Russell, 2005; Schönborn and Anderson, 2010), a profound analysis of their drawing processes is still lacking. Supporting students in drawing, therefore, requires a closer look at how students are sequentially generating their drawings. Recording and analyzing the eye-movement trajectories of students can help to characterize to what extent students use and perceive different drawing elements to generate resonance structures and to determine how different gaze patterns may relate to the productivity of the generated drawings. Specifically, by examining students' drawing processes quantitatively and qualitatively, we seek to answer the following research questions (RQ) in this exploratory study:

1 What drawing elements do students connect when constructing productive or unproductive resonance structures (i.e., in terms of transitions and relations of the Areas of Interest)?
2 What structural features do students attend to in terms of attention distribution when translating one resonance structure into another and how is it related to students' drawing moves?
3 How does a student's approach to visual selection to construct resonance structures relate to the productivity of the drawings?

With regard to the first research question, we assume that students with unproductive drawings might exhibit a more varied search behavior, thus, cognitively connect more drawing elements in the construction process, including the use of and the transition between unrelated drawing or task elements. This hypothesis is supported by previous eye-tracking research across different STEM education disciplines as it has been repeatedly shown that successful problem-solvers fixate more on relevant aspects of a representation and generally show a more focused behavior, whereas unsuccessful problem-solvers exhibit a more distributed visual behavior including more fixations on irrelevant aspects of a given representation (Tang and Pienta, 2012; Hejnová and Kekule, 2018; Havelková and Gołębiowska, 2019). In accordance to that, it has been demonstrated that inexperienced individuals more often attend to salient information (i.e., stimulus-driven information selection) that may be irrelevant for task performance. Experienced individuals, on the other hand, efficiently attend to information knowing which information is important for task performance (goal-driven selection) (Jarodzka et al., 2010). Based on the consideration of visual search mechanisms, we assume for the second research question that the construction of unproductive drawings may stem from difficulties to select more specific task-relevant information, such as identifying interacting structural features in a resonance structure, and may result in an overreliance on single, salient structural features. Finally, we assume for the third research question that students who explicitly apply conceptual knowledge to the construction of resonance structures, such as the application of rules or the inference of implicit structural properties, more often construct productive drawings than students whose drawing approach is characterized by a mere rearrangement of surface structural features (Graulich et al., 2019).

Materials and methods

Context and participants

This research study was conducted at a German university in summer 2021. Students were recruited on a voluntary basis via e-mail and in-class announcements in the Organic Chemistry 1 (OC1) course and were given 20 euros as compensation for their time. A total of 21 students agreed to participate in this study. The participants' ages ranged from 19 to 34; nine of them identified as male and twelve as female. One student was excluded from the analysis due to measurement errors, so the number of further analyzed participants was reduced to 20. All students were majoring in chemistry (i.e., they were chemistry, food chemistry, and chemistry teacher students) and were beginners in organic chemistry, i.e., all of them had taken OC1 as the first lecture in organic chemistry either in the second or fourth semester of their studies. Normally, this course consists of a weekly lecture (3 h) and weekly tutorial sections (1.5 h). Due to the pandemic situation, in summer 2021, a flipped format was adopted in which students watched tutorial videos and read material (e.g., book chapters) prior to solving content-related tasks and discussing questions in online tutorial sections. The OC1 course provides basic knowledge of organic chemistry, covers the reactivities of functional groups, deals with structure-property relationships and discusses typical reaction mechanisms (e.g., radical substitution, electrophilic addition, nucleophilic substitution, and carbonyl reactions). The study took place near the end of the course to assure that the students were familiar with the reactions used in this study, the resonance concept, and the construction of resonance structures.

All students who volunteered were informed about their rights and data handling beforehand; informed consent was obtained from all participants. Institutional Review Board approval was not required for this study. Nevertheless, the study followed ethical guidelines and it was clarified to the students that they had the opportunity to opt out at any time. All students gave their written consent for the collected data (i.e., their recordings and scans of their worksheets) being analyzed and published by the research team. In this study, participants were assigned pseudonyms and no identifying information was recorded or scanned to allow participants to re-identify. As the interviews were conducted in German, students' interview excerpts were translated into English for this publication.

Data collection

Study design

The study followed a qualitative approach and used a semi-structured interview to explore both students' reasoning with drawings and students' drawing processes of structural formulas in organic chemistry (**Figure 1**). Before starting the interview, the interviewer explained the interview procedure and briefed the students that the topic of the study would be the resonance concept. In the first section, general questions about the resonance concept (e.g., "When does resonance occur?") have been asked to refresh the students' minds and to gather information on students' abilities to draw resonance structures as well as to infer chemical information from the structural formulas (e.g., "What impact does resonance have on the energy of a structure?"). In the subsequent main section, the students were prompted to solve three organic case comparison tasks (Graulich and Schween, 2018) requiring resonance structures. We used case comparisons to elicit students' drawing process as they necessitate drawing resonance structures in order to solve the given problems and estimate the differences between the two given reactions. All reactions were covered in the OC1 lecture. In the last section, problems with regard to the completion of the tasks as well as students' needs for these types of tasks were addressed in a general reflection.

The subsequent analysis focuses on the third case comparison task as it allows the direct comparison of students' drawings. It describes the mechanistic leaving group departure step of a nucleophilic substitution reaction and asks the students to determine which of the two reactions would form the most stable product. As shown in **Figure 1**, bromide leaves as leaving group under the formation of a carbocation in both reactions. In each case, the substrate is a primary alkyl substrate which

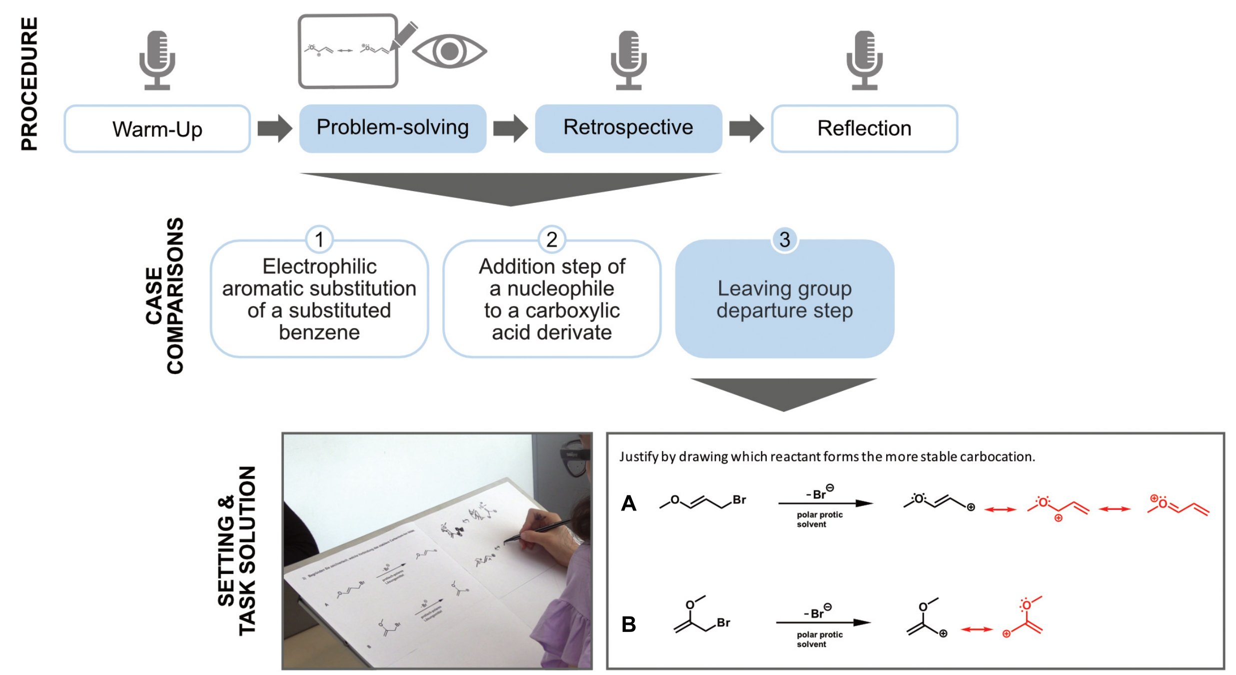


FIGURE 1
Study design and task solution for the utilized case comparison. Blue boxes indicate the results highlighted in this report.

enables stabilization via resonance. Only the position of the methoxy group differentiates the substrates. To determine the most stable carbocation, one has to draw and evaluate the electron density distribution in the resonance structures of both products. While there are two (productive) resonance structures in B, there are three in A. It follows that the positive charge can be better distributed across the whole molecule in A. This is responsible for the lower potential energy and, consequently, for the higher stability of product A.

Qualitative interview

A qualitative semi-structured interview guided the investigation of students' drawing and reasoning process. The interviews were conducted face-to-face between one participant and the interviewer, lasted between 83 and 140 min and were audio- and video-recorded. All materials were administered in a pencil-paper setting and the participants were encouraged to solve all the tasks freely, thus, they could write or draw as much as they considered necessary. To capture students' eye movements in their natural problem-solving behavior, they were not disturbed or had to explain their approach during the completion of the case comparison tasks. However, they were allowed to think aloud if they wanted. After completing each case comparison task, a subsequent retrospective interview focused on students' rationale for their drawings and their task specific problem-solving process. For instance, the students were asked to justify their final choice and to describe their drawing strategies (e.g., "How did you get to this structure?").

Eye-tracking

A mobile eye-tracker (Tobii Pro glasses 3, 50 Hz) served as a tool to capture the participants' eye-movements while drawing during the problem-solving process. To optimize the collection of the eye-tracking data, a drafting table was used and each task was presented on the upper left side of a 42.0 × 59.4 cm (DIN A2) sheet of paper (Figure 1). The eye-tracking glasses were calibrated and validated individually for each participant prior to solving the tasks. In case of nearsightedness and farsightedness, suitable corrective lenses were used. A vision test also validated their fit. All gaze samples ranged above 79% (average 91.7%).

Data analysis

Eye-tracking data

Data preparation

First, the gaze data of all participants' recordings during the problem-solving phase, i.e., until the students gave their answers prior to subsequent possible refinements of the drawings, were manually mapped by a trained student research assistant using the software *Tobii Pro Lab*. The first author double-checked the mapped gaze points to check accuracy. Second, for every participant, the Times of Interest (TOI) were defined for each drawing event (i.e., the time sequence until completion of a single student-generated resonance structure). The duration of the overall drawing process for the task varied for each participant ranging from 1.02 to 12.25 min (average 4.06 min). The construction of the single resonance structures took the participants from 6 s to 5.26 min (average 51 s).

Third, the Areas of Interest (AOI) were defined for each drawing event (i.e., every constructed resonance structure until the final decision of the student) for each student. For the subsequent analysis, the AOIs were set on different grain sizes. An AOI was defined for every complete drawing and for smaller parts of the respective molecules, always maintaining the main features of the structures (i.e., the methoxy group, the double bond, and the positive charge) (**Figure 2A**). All eye-tracking analyses were conducted with the software *Tobii Pro Lab* and *RStudio*.

Furthermore, the students' drawings were classified as productive, unproductive, and auxiliary drawings. While productive drawings encompass all correct resonance structures that help answer the given task, unproductive drawings are either incorrect resonance structures (e.g., violating the octet rule) or drawings that are technically correct, respecting the octet rule, but are, nevertheless, insignificant for this task due to the number of formal charges. The auxiliary drawings are either structures that the students copied from the given task or structures or texts that the students wrote for themselves and were, thus, not directly related to the construction of the resonance structures. They were thus not taken into consideration in the following analysis. **Figure 3** provides an overview of the different drawing categories. Next, the data were analyzed in several steps using a combination of quantitative and qualitative approaches.

Examining the connection of drawing elements during the drawing process

To determine how students used the different drawing elements to construct subsequent resonance structures, the analysis of students' gaze behavior was twofold, looking at the structures' co-occurrence during the entire drawing process and the transitions between these structures (RQ 1). In doing so, we grouped all AOIs for each drawing event at a broader grain size into four categories to enable the comparison of the

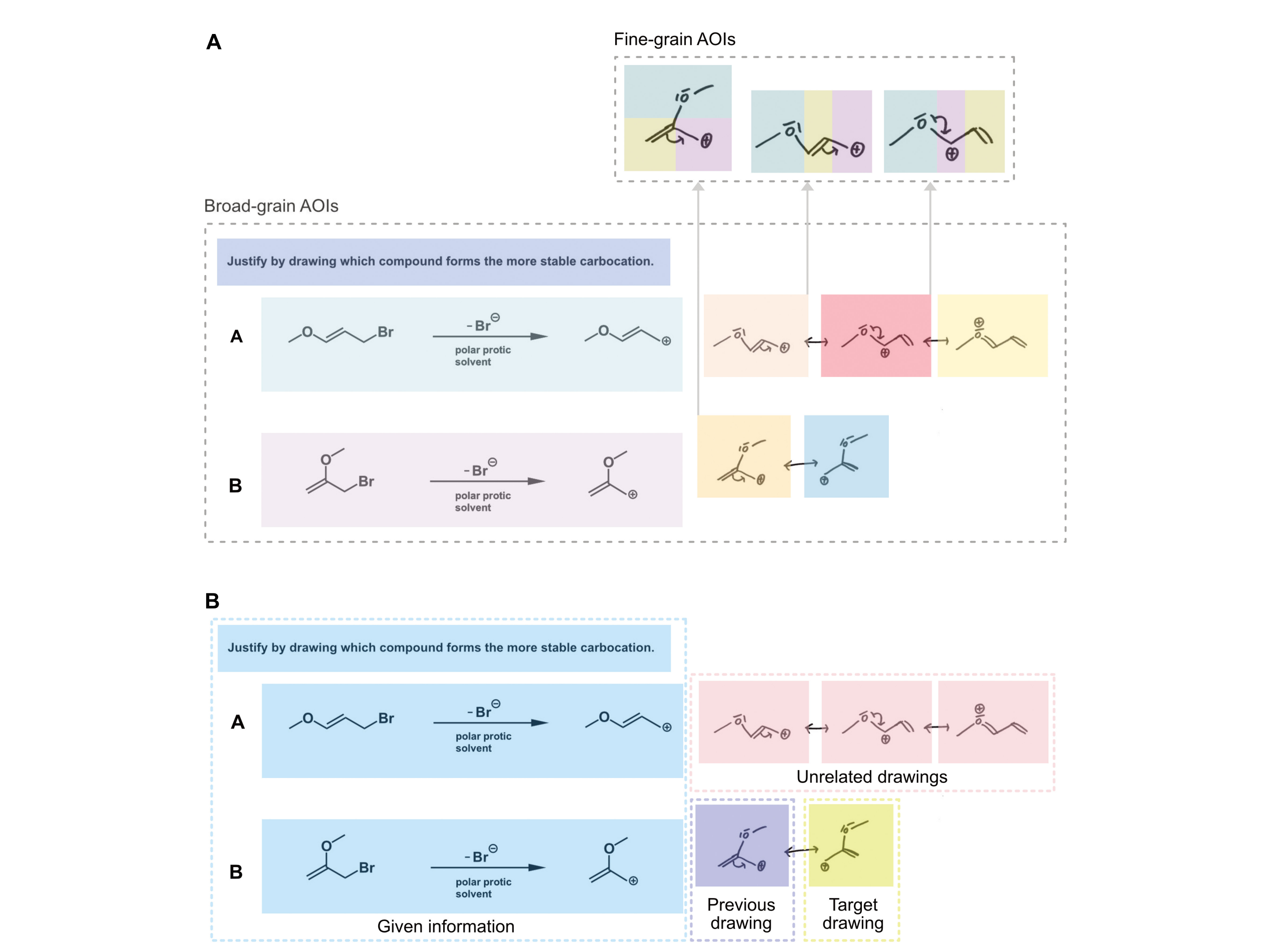


FIGURE 2
(A) Definition of the AOIs of the drawing elements at a broader and finer grain size exemplified by a student solution and **(B)** exemplary categorization of the AOIs in dependence of their function (target drawing, previous drawing, given information, unrelated drawings).

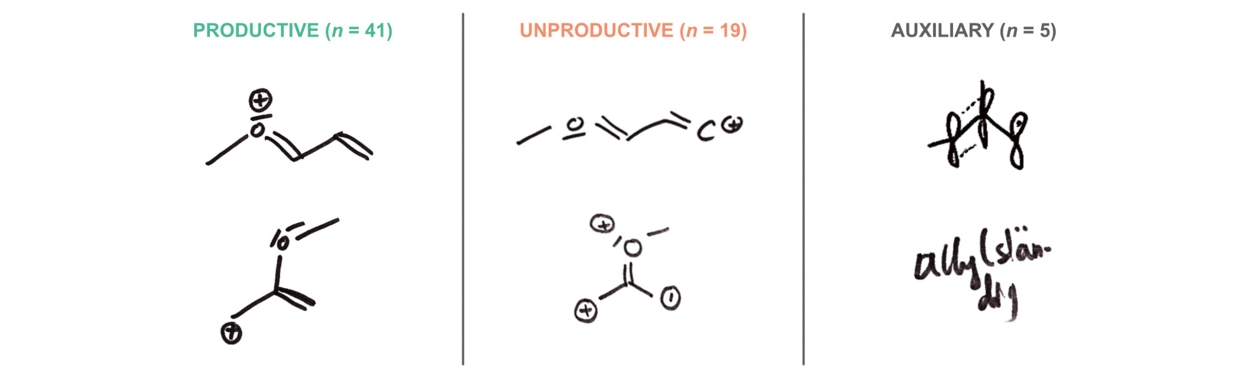


FIGURE 3
Classification categories for the resulting drawings with student examples.

various construction processes of every student: target drawing (T), previous drawing (P), unrelated drawings (U), and given information (G) (cf. **Figure 2B**). While the target drawing is the resonance structure of interest, the previous drawing represents the drawing from which the target drawing results. The given information consists of all the structures and text available to the students in the task prompt. The transformation of one structure into another is carried out at the local structure, i.e., apart from the previous drawing, no other information is necessary in order to construct the target drawing. Therefore, we labeled all the structures and texts that the students might have noted but that are *per se* not necessary for constructing the target drawing (e.g., resonance structures of the product in A when the participant is constructing the resonance structures in B as well as additional explanatory text) as unrelated drawings. Although the given information is neither required for the construction of resonance structures and thus also represents unrelated information, we maintained this category to investigate whether there are tendencies of revisits as anchors when constructing a productive or unproductive drawing.

First, we conducted an Epistemic network analysis (ENA) to examine the co-occurrence of the various drawing elements during the construction of resonance structures and, thus, to explore different gaze patterns depending on the resulting productivity of the drawing. The ENA is an analytical method from the field of learning analytics, which has been used in numerous contexts, including eye-tracking analyses (Andrist et al., 2015; Shaffer et al., 2016; Bruckner et al., 2020). This method can be used to identify and quantify the structure of connections among coded data elements in any system by representing their associations in networks, i.e., the mere presence of isolated elements is not as important as their interrelations (Shaffer et al., 2016). Thus, the ENA allows the characterization of even complex and dynamic relationships (e.g., patterns of association within discourse or gaze behavior) by illustrating both the structure and the strength of connections in both single networks (e.g., in terms of their plotted point position) and network difference graphs which illustrate the differences between two networks by subtraction (Shaffer et al., 2016; Shaffer and Ruis, 2017). While the nodes in the network correspond to the analyzed codes and appear in every modeled network in exactly the same position, the edge width between the nodes reflects the relative frequency of co-occurrences between two codes. In order to model the connections, ENA uses a singular-value decomposition which performs dimensional reduction on a high-dimensional space, producing fewer dimensions that capture the maximum variance in the data (Shaffer et al., 2016).

In our analysis, we used the *ENA Web Tool* (version 1.7.0) to compare the weighted epistemic networks of the construction processes of productive and unproductive drawings on the basis of students' collapsed AOI hit sequences during the construction of resonance structures. To this end, each drawing served as a *unit*, i.e., the piece of data for which the ENA constructs networks. *Conversations* indicate how to segment the data for analysis in terms of their relation to one another, i.e., *units* not in the same *conversation* are not related to one another in the network model. As we aimed at characterizing the drawing process of each resonance structure individually, again, we chose as *conversation* every drawing and chose the mode *whole conversation* so that the ENA modeled connections across the entire *conversation*. In the ENA, the nodes of the network represent the *codes*. In order to be able to compare the various drawing processes, our *codes* comprised the previously defined categories "given information," "unrelated drawings," "previous drawing," and "target drawing" for every drawing event. Finally, as *comparison* served the productivity of the drawings, i.e., we differentiated between productive and unproductive drawings.

In a second step, the same AOI hit sequences used in the ENA beforehand were analyzed to determine the transitions between individual pairs of AOIs that the students made to construct resonance structures in order to reveal how students integrated the various drawing elements (i.e., the defined four drawing categories) when constructing structural formulas

(Schmidt-Weigand et al., 2010; Johnson and Mayer, 2012). For that purpose, we used the *GrpString R package* (Tang et al., 2018) to calculate the transition matrix, the transition entropy and the total amount of transitions across the groups of unproductive and productive drawings, regardless of whether the drawings resulted from a previously productive or unproductive drawing as no discriminating differences could be found in a finer division. The transition entropy indicates the diversity of the transitions within a string or different string groups. While a higher entropy reflects more evenly distributed transitions, a lower entropy measure indicates a more biased distribution of transitions, mainly reflecting transitions between fewer AOIs (Tang et al., 2018). The transition matrices were also explored for statistical significance using the Mann-Whitney *U* test. A non-parametric test was chosen because a Shapiro-Wilk test ($p < 0.05$) indicated that not all of the data to be used for comparison are normally distributed (Shapiro and Wilk, 1965).

Analyzing the attention distribution while constructing resonance structures

To gain deeper insight into the translation process of single resonance structures, the analysis further concentrated on the connection of each drawing pair, i.e., the previous and target resonance structure (RQ 2). As the attention distribution may indicate the ascribed importance of a representation (Cullipher and VandenPlas, 2018), we examined the total fixation duration of the AOIs of the previous drawing to which students attended when constructing the target resonance structure. In doing this, we used the AOIs at the smaller grain size encompassing the different structural features of the respective previous resonance structure (cf. **Figure 2A**). With this data, we then calculated the ratio of attention distribution on each structural feature of the previous resonance structure. Again, we differentiated between productive and unproductive drawings for all drawings to examine possible differences in attention distribution.

Qualitative data

Determining the relationship between drawing moves and attention distribution while constructing resonance structures

To relate and compare the attention distribution to students' drawing moves for possible relationships (RQ 2), students' drawing moves made to get from one structure to the next one were inductively analyzed (i.e., the delocalization of electrons and change of the structural features). Different drawing moves could be identified which can be subsumed either as single drawing move (e.g., delocalizing one electron pair or charge) or as multiple drawing moves (e.g., delocalizing several electron pairs) (**Figure 4**).

The first author coded all drawings. Additionally, a trained student research assistant coded the entire data independently with the code book. A kappa coefficient κ_n of 0.87 (Brennan and Prediger, 1981) was calculated, showing high agreement and reliability for the coding rubrics (Kuckartz and Rädiker, 2019). Any ambiguities were discussed and resolved to reach a final agreement of 100%.

Characterizing students' approaches of visual selection

All audio recordings of the interviews were transcribed verbatim and implemented into the coding software *MAXQDA* for subsequent qualitative content analysis (Saldaña, 2016). To examine students' approaches of visual selection when constructing resonance structures and, consequently, triangulate the eye-tracking data findings, our qualitative

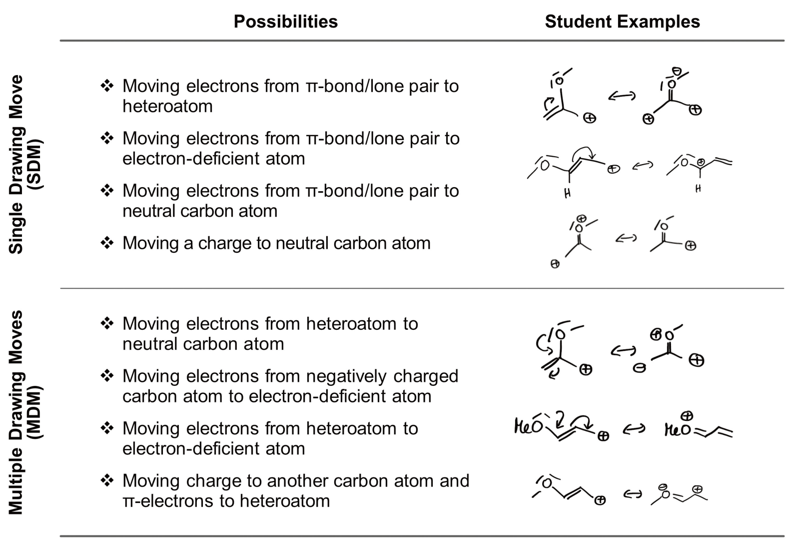


FIGURE 4
Coding scheme for drawing moves with student examples.

analysis centered on how students verbally described their drawing process (RQ 3). The analysis was informed by the theory of visual selection described in section "Mechanisms of visual selection" (Theeuwes, 2019). However, we could not adopt the three different mechanisms of visual selection as codes due to different reasons. First, the problem-solving process in this study, as a whole, is clearly a top–down process because the participants knew that they had to use the resonance concept to solve the task and, consequently, to draw resonance structures by moving electrons within structural formulas. Therefore, the task demands by definition a goal-driven approach to construct the resonance structures. Second, it is not possible to properly distinguish between a stimulus-driven, history-driven, and goal-driven approach in many chemistry contexts, because reasoning and sense-making often depend on the interplay of deriving and perceiving explicit and implicit structural features and properties (Graulich et al., 2019). Students can exhibit both a bottom–up and top–down approach. Therefore, we adapted the aforementioned mechanisms into two categories which we applied as a lens to analyze students' descriptions of their drawing approach: a knowledge-driven approach and a structure-driven approach. The code "structure-driven" was ascribed whenever the students described their drawing process by only referring to and mentioning the explicitly drawn structural features. Thus, it can be considered a bottom–up process resembling the stimulus-driven approach. The code "knowledge-driven" was given when the students explicitly used and verbalized their knowledge to construct the respective resonance structure. This may include different kinds of knowledge, e.g., experiences or concept knowledge such as the reference to (implicit) chemical concepts (e.g., stability), or rules (e.g., the octet rule). Therefore, this approach can be considered a top-down process that resembles the goal-driven and history-driven approach.

Moreover, we characterized the flexibility of students' approaches. While a "centered" approach applies when students were focusing on one structural feature (e.g., the methoxy group of the molecule), students with a "variable" approach described their drawing process by taking into account the entire structure or at least several interacting structural sections of the starting

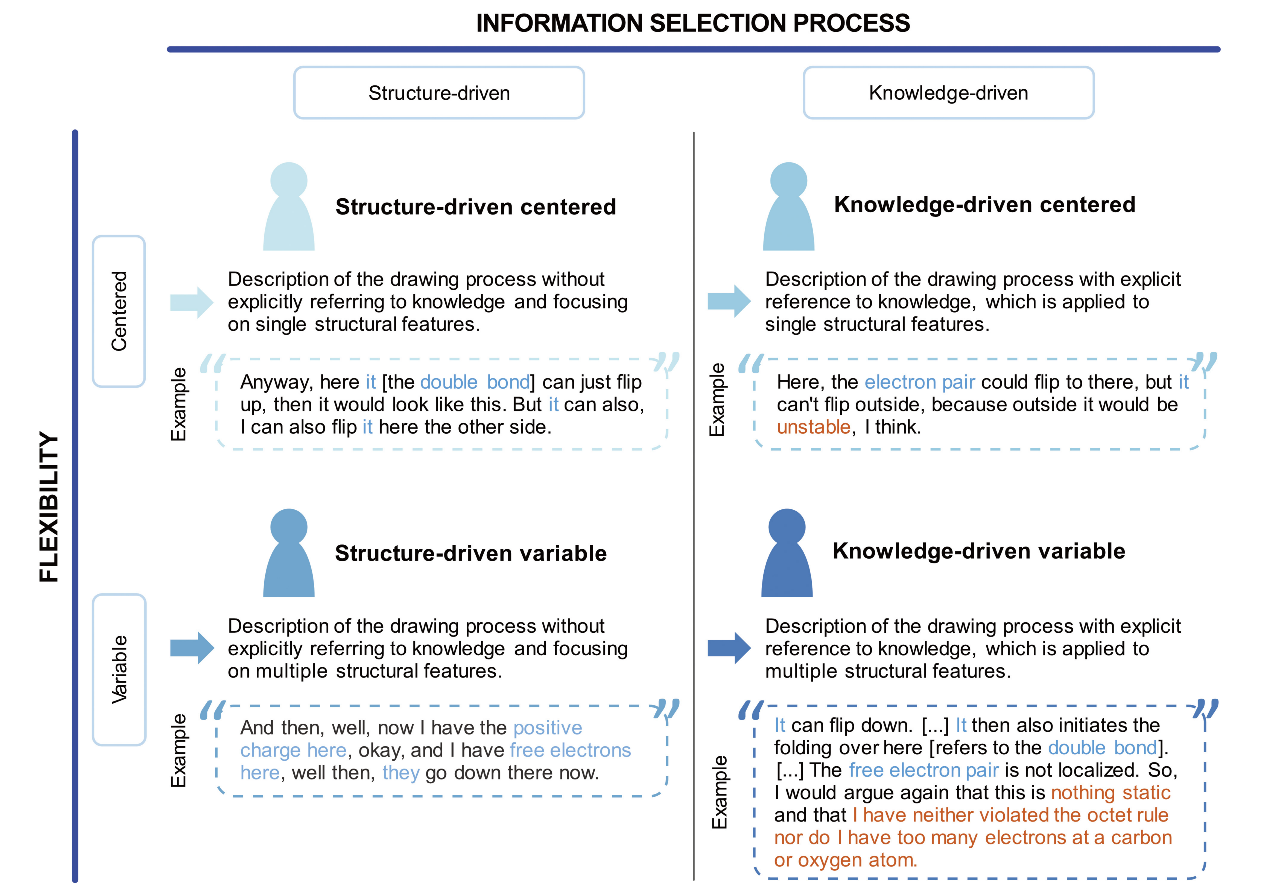


FIGURE 5
Coding scheme for students' visual selection approaches with definitions and student examples. The highlighted blue text in the quotes indicates the structural features the students referred to, while the orange text highlights the verbalized knowledge (e.g., rules and chemical concepts).

molecule. Figure 5 provides the definitions of the resulting four codes and illustrates them by giving student examples for each coding category.

The first author coded the entire data set. During the data analysis, the authors regularly discussed and optimized the coding scheme to ensure that coding decisions faithfully represented the data. For interrater reliability, the second author coded a random sample of 20% of the data independently. A kappa coefficient κ_n of 0.94 (Brennan and Prediger, 1981) was calculated, indicating high agreement and reliability for the coding rubrics (Kuckartz and Rädiker, 2019). Any ambiguities were discussed and resolved. In the end, a 100% agreement between the two authors was reached.

Results and discussion

Of all 60 student-generated resonance structures, 41 drawings were productive, whereas 19 were unproductive, often resulting from previously productive drawings or building a sequence of unproductive resonance structures. Therefore, it is of interest to characterize what distinguishes the construction process of productive from unproductive drawings. Using a Mann–Whitney *U* test, we compared the construction of productive and unproductive drawings on the basis of different eye-tracking data (Table 1). Addressing our research questions, the subsequent sections present and discuss the main findings.

TABLE 1 Results of the Mann–Whitney *U* test for different eye-tracking data comparing the construction of productive and unproductive drawings.

Variable	$Mdn_{pr.}$	$Mdn_{unpr.}$	U	z	p^a	r^b
Connection of AOIs (ENA)						
x-axis/MR1	-0.12	0.13	497.50	2.169	0.09	0.28
y-axis/SVD2	-0.32	-0.37	359.20	−0.620	0.64	0.08
Transitions						
Absolute number	15	17	378.5	−0.175	0.861	0.02
Entropy	1.96	2	342.5	−0.747	0.455	0.10
Transition types						
GU	0	0	384	−0.141	0.888	0.02
GP	0.053	0.067	363.5	−0.439	0.661	0.06
GT	0	0	359	−0.78	0.435	0.10
UG	0	0	386.5	−0.066	0.947	0.01
UP	0	0.029	367	−0.388	0.698	0.05
UT	0	0	252.5	−2.797	**0.005**	0.36
PG	0	0	367.5	−0.377	0.706	0.05
PU	0	0	371.5	−0.377	0.706	0.05
PT	0.33	0.24	262	−2.028	**0.043**	0.26
TG	0	0	378	−0.328	0.743	0.04
TU	0	0	266.5	−2.637	**0.008**	0.34
TP	0.29	0.16	249	−2.236	**0.025**	0.29
Fixation duration rate						
Relevant features	0.78	0.50	188	−3.212	**0.001**	0.41
Unrelated features	0.22	0.50	591	3.212	**0.001**	0.41

Significance level of 0.05 and confidence interval of 95%; pr., productive; unpr., unproductive; $N = 60$; [a]Significant *p*-values (<0.05) are displayed in bold. [b]Calculated as indicated by Rosenthal (1984).

RQ 1: What drawing elements do students connect when constructing productive or unproductive resonance structures?

On a global level, students exhibit a similar gaze behavior when using information from different drawing elements to construct productive or unproductive resonance structures. In fact, as the mean epistemic networks of productive and unproductive drawings illustrate (Figures 6A,B), the co-occurrences of the different drawing categories (i.e., target drawing, previous drawing, unrelated drawings, and given information) appear with a similar density and show, in general, no considerable differences concerning the type and frequency of connections.

Consistent with this observation, a Mann–Whitney *U* test showed that neither along the *x*-axis (MR1) nor the *y*-axis (SVD2) productive drawings differ significantly from unproductive drawings (Table 1). From a qualitative perspective, the association between the previous drawing and the target drawing dominates in both groups. Comparing the different networks for productive and unproductive drawings, represented by the position of the respective network graph node (Figure 6C), no distinctive tendency concerning the patterns of co-occurrence of the drawing elements can be inferred. This result indicates that the construction of an (un-)productive resonance structure does, thus, not globally depend on the resulting network, i.e., the amount and density of associations of the different drawing elements. Hence, our hypothesis of unproductive drawings resulting from the connection of more drawing elements cannot be confirmed. However, the subtracted mean network (Figure 6C) shows that there exist differences concerning single connections. Students with productive drawings (green color) exhibit a stronger association of the previous and target drawing, whereas students with unproductive drawings (orange color) show more associations of the unrelated drawings with both the target drawing and the previous drawing. This indicates that using unrelated information seems to play a bigger role when constructing unproductive structural formulas. The subtracted mean network also illustrates that in both groups, the associations between the other drawing elements are altogether

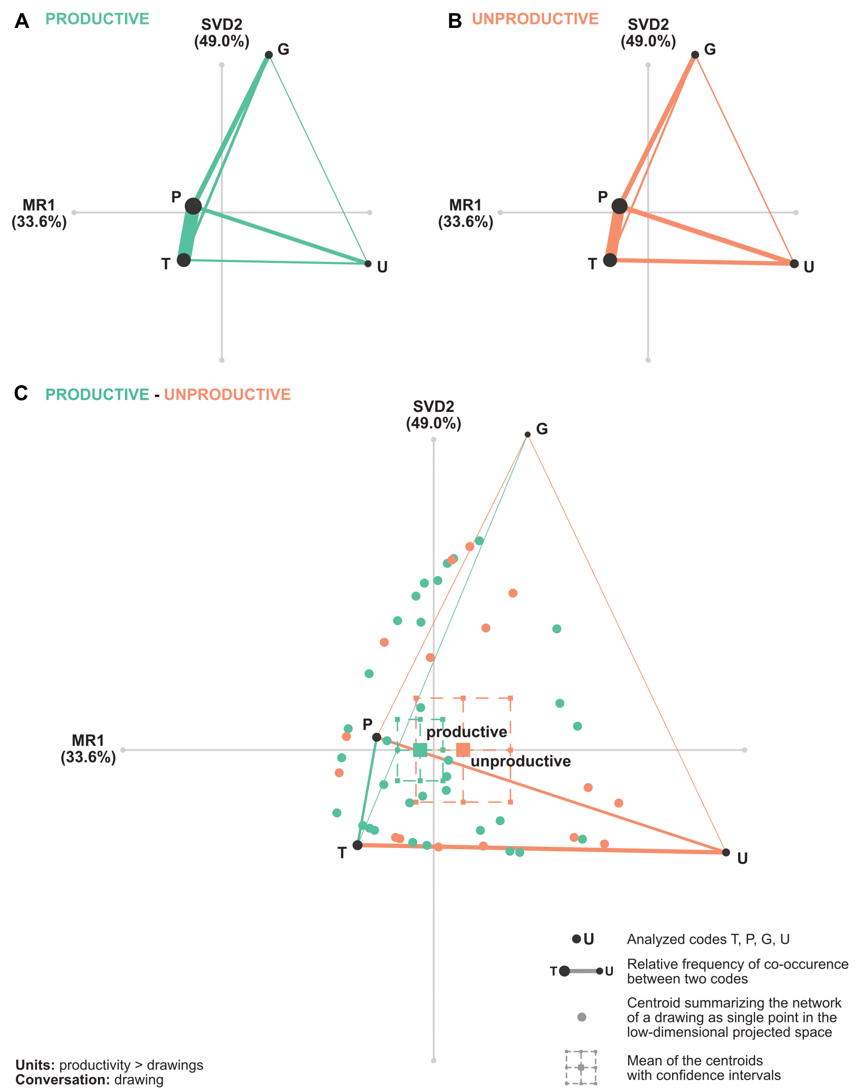


FIGURE 6
Epistemic network for **(A)** productive drawings, **(B)** unproductive drawings, and **(C)** the subtracted epistemic network of productive and unproductive drawings using the collapsed AOI hit sequences during the construction process of every drawing (*N* = 60). The drawing categories (T, target drawing; P, previous drawing; G, given information; U, unrelated drawings) serve as codes.

scarce and that the given information rarely co-occurs with the other drawing elements.

Although the ENA models the gaze-pattern networks of the different construction processes of (un-)productive drawings by depicting the weighted frequency of co-occurrences of the different drawing elements, it provides no insight into the direction of linkage, i.e., transitions, between those elements (e.g., in terms of backtracking). The ENA only depicts how often drawing elements co-occur altogether. However, transitions play a crucial role when processing information (Schmidt-Weigand et al., 2010). Thus, analyzing the linkage direction may also reveal differences regarding the direct connection of drawing elements, such as integrating information to construct productive and unproductive drawings.

A look at the different transitions during the drawing process of the resonance structures reveals that in general, in accordance with the ENA, the transition entropy for productive or unproductive resonance structures does not differ significantly (Table 1), thus, the gaze transition distribution is similar for both groups. Neither does the absolute number of transitions differ significantly. However, as listed in Table 1, there are significant differences between the types of transitions during the drawing process of a productive and unproductive drawing regarding the transitions to the target drawing.

Figure 7 depicts the subtracted transition matrix of productive and unproductive drawings with green cells indicating transitions more characteristic for productive drawings. In contrast, red cells show transitions rather occurring

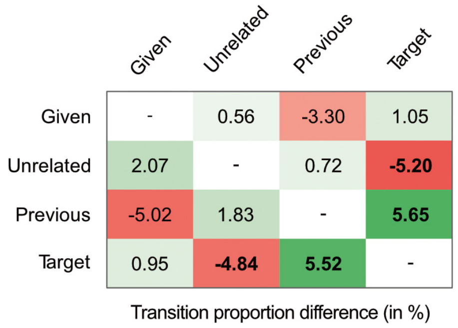


FIGURE 7
Subtracted transition matrix indicating characteristic transitions regarding the construction of productive (green cells) and unproductive drawings (red cells); significant differences in transition types are displayed in bold (p-value of < 0.05).

with unproductive drawings. According to **Table 1**, it becomes evident that the drawing process of productive resonance structures encompasses significantly more transitions between the previous drawing and the target drawing ($p = 0.043$, $r = 0.26$) and vice versa ($p = 0.025$, $r = 0.29$). In comparison, the drawing process of unproductive drawings comprises more transitions between the target drawing and unrelated drawings ($p = 0.008$, $r = 0.34$) and vice versa ($p = 0.005$, $r = 0.36$). **Figure 7** shows as well that students with unproductive drawings connect the given information with the previous drawings more often than students who construct productive drawings. However, the difference of this transition type is statistically not significant. Given that transitions play a crucial role in the integration and processing of information (Schmidt-Weigand et al., 2010; Johnson and Mayer, 2012), an increased integration of unrelated information can impede the construction of valid resonance structures. In fact, integrating more information in the target drawing by focusing on diverse, less relevant, drawing elements, can indicate searching processes and uncertainty. That said, it is possible that learners try to align and transfer information from previous drawing processes or search for anchor points. This assumption is in line with previous research on problem-solving and the comprehension of visualizations (e.g., Gegenfurtner et al., 2011; Tang et al., 2014). For instance, Holmqvist et al. (2011) showed in the context of a mathematical problem-solving task that participants with low ability tended to scan all AOIs in the tasks while high-ability participants exhibited a more focused behavior. Existing research in STEM education also supports the finding that students' drawing process of productive drawings is characterized by more direct transitions from previous to target drawing. Therefore, multiple studies show that high-performing participants (e.g., experts) more effectively process information, e.g., by transitioning between relevant parts of tasks (Baluyut and Holme, 2019; Connor et al., 2021), by efficiently attending to task-relevant features (Jarodzka et al., 2010; Tang and Pienta, 2012; Topczewski et al., 2016; Hejnová and Kekule, 2018; Havelková and Gołębiowska, 2019), or by exhibiting a more focused searching behavior when problem-solving (Holmqvist et al., 2011; Rodemer et al., 2020).

However, despite the group-specific differences in transitions, the connection between the previous drawing and the target drawing overall plays a major role in students' construction process of resonance structures. This suggests that the students in this study already possess a distinct representational fluency concerning the direct translation from one representation into another. With regard to representational competence, as described by Kozma and Russell, students with productive and unproductive representations show similar abilities to identify and select structures necessary to transform a resonance structure into another (Lesh et al., 1987; Kozma and Russell, 1997, 2005).

RQ 2: What structural features do students attend to in terms of attention distribution when translating one resonance structure into another and how is it related to students' drawing moves?

Since all students mainly connected the previous and the target drawing when constructing productive and unproductive resonance structures, it is of interest to investigate students' attention distribution when translating one structure into another, and thus, to consider their fixation duration on the different parts of the starting molecule, the previous drawing, in order to characterize their decoding behavior (R-M factor) (Schönborn and Anderson, 2008, 2010).

Figure 8 illustrates the fixation duration rate for the construction of productive and unproductive resonance structures on the basis of productive, initial drawings (57 of 60 drawings). Due to their very specific and individual character, **Figure 8** does not comprise the remaining three structures resulting from unproductive drawings. The frames of the respective pie charts indicate the adjacent parts of the molecular structure which are relevant for translating this structure into the corresponding productive structure. As **Figure 8** illustrates, the fixation duration rate differs with regard to the constructed structures, showing an emerging attention distribution difference on different parts of the molecules when constructing productive or unproductive drawings. As indicated in **Table 1**, the fixation duration rate of relevant and unrelated structural features for productive and unproductive drawings differ significantly with medium effect size (relevant features: $p = 0.001$, $r = 0.41$; unrelated features: $p = 0.001$, $r = 0.41$). For the productive resonance structures, a slight trend emerges toward an increased attention to the relevant structural features for the resulting resonance structure. If we consider the productive subsequent resonance structures A1

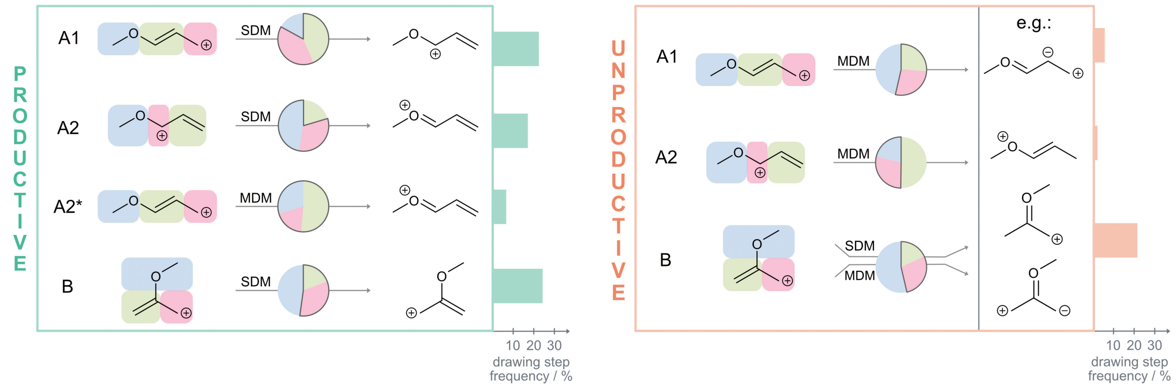


FIGURE 8
Relative drawing step frequency (N = 57) and fixation duration rates (pie charts) on the structural features of the initial structures (blue = methoxy group, green = double bond, pink = positive charge) when constructing a subsequent productive (green frame) or unproductive (orange frame) structure. The relevant, necessary structural features for the construction of the following structure are highlighted by framing. Drawing moves are indicated as single drawing move (SDM) and multiple drawing moves (MDM).

and A2, students who construct these structural formulas on average focused more on the interrelated structural features necessary for the construction of the respective structure, i.e., on the positive charge and the double bond in the first structure (83% of the attention distribution), and the positive charge and the methoxy group in the second structure (80% of the overall attention distribution). To draw the productive structure A2* on the basis of the given, initial product in A, thus, skipping the second structure, students have to consider every part of the structural formula as electrons are delocalized throughout the whole molecule. However, it becomes apparent that students' attention was guided by the double bond and the methoxy group (altogether 81%) and less by the positive charge (19%) that, however, represents a productive starting point for the drawing process.

Figure 8 further illustrates that few students constructed unproductive drawings in reaction A, while most unproductive drawings resulted when constructing resonance structures in reaction B. In general, it can be derived from **Figure 8** that unproductive drawings stem from a higher attention to unrelated structural features. For example, consider the attention distribution of the unproductive structure A1. In contrast to the productive structures A1 and A2, students shifted their attention more to the methoxy group (46%). In the unproductive structure A2, students payed their attention twice as often to the unrelated double bond (50%) whereas the structural features that are relevant for the construction of a productive structure (i.e., the methoxy group and the positive charge), were considered less (50%) compared to when students constructed a productive structure based on the same initial drawing. This trend of overly considering unrelated structural features while paying less attention to relevant, interacting structural features also applies to the unproductive resonance structures in B as students paid much attention to the methoxy group. However, here, the fixation durations rates do not differ as much from those that students exhibit when generating productive resonance structures in B. Regarding representational competence, it can be derived from these findings that although students are able to identify and connect relevant structures in order to construct subsequent resonance structures (Kozma and Russell, 2005), they differ in the decoding and interpretation of the representations by paying attention to different structural features. Consequently, they may perceive the external features of a structure differently (R-M factor), eventually resulting in unproductive drawings (Schönborn and Anderson, 2008, 2010).

Besides the attention distributions, **Figure 8** depicts the drawing moves the students made to construct their drawings. In this task, productive drawings often stem from a single drawing move where electrons are delocalized toward an electron-deficient atom. As **Figure 8** shows, students predominantly focus thereby on two adjacent, interrelated structural features, that is, the positive charge and the adjacent double bond or methoxy group. In contrast to that, students who perform multiple drawing moves construct either unproductive resonance structures, or draw a productive resonance structure but tend to skip a structure, which, however, would be necessary to answer the given task. Altogether, 13 of the 19 unproductive drawings are based on multiple drawing moves which may either involve arbitrarily moving charges, moving the lone pairs of the heteroatom throughout the whole molecule or moving electrons to an electron-rich atom (cf. **Figure 4** for examples).

While there is a clear difference with respect to the fixation duration rate in the structures of reaction A, the gaze proportion on structures in reaction B does not differ much. Hence, our hypothesis for the second research question, that the construction of unproductive drawings results from difficulties in the selection of relevant information and, eventually, an overreliance on single, salient structural features, can only be partly confirmed. This lack of difference for reaction B may be due to different reasons. First, the different geometries of the molecules can be responsible for a different decoding behavior. While the molecules in reaction A are linear, the molecule in B is more branched, thus, spatially more complex. Following this, the decoding of the molecule and the identification of interacting structural features could have been cognitively more demanding and may have led to longer fixations of the methoxy group in order to decide upon its possible influence in the structure. In fact, a longer fixation on a structural feature correlates with longer mental processing of this information (Just and Carpenter, 1980), which may result from difficulties in interpretation or due to the perceived importance of the feature because of its salience (Cullipher et al., 2018). As students in both groups fixated much on the methoxy group, it may be that this structural feature was more difficult to interpret than the double bond, since a functional group containing an oxygen atom can have an electron-donating or electron-withdrawing function depending on its connectivity. Therefore, students may have tried to connect this feature with prior knowledge such as rules concerning its electronic effects (R-C factor) (Schönborn and Anderson, 2008, 2010). Our findings of students' varying decoding and use of structural features when constructing resonance structures align with existing research on representational competence in chemistry. While it has been shown that students often exhibit difficulties regarding the comprehension and interpretation of representations (e.g., Keig and Rubba, 1993; Kozma and Russell, 1997; DeFever et al., 2015), more specifically, Olimpo et al. (2015) have shown that the translation of Newman projections and Dash-Wedge representations is easier for students when dealing with less complex molecules. Moreover, using eye-tracking, in Baluyut and Holme's (2019) study it became evident that the visual complexity of particulate nature of matter diagrams impacted students' viewing behavior as low-performing students fixated on more features of a task when working with such representations. Rodemer et al. (2020), on the other hand, showed that visual characteristics seem to influence students' visual processing and problem-solving in organic chemistry, with more factors embedded in a case comparison leading to more transitions between representations.

Besides the higher spatial complexity, familiarity may also play a crucial role regarding students' gaze behavior in case of the linear structures in reaction A. Both the allylic position of the carbocation and the linearity of the molecule are familiar to the students from the lecture. Therefore, it is possible that students felt more confident in decoding the structures and, consequently, could determine relevant, interacting structural features more easily, narrowing their gaze distribution on specific parts of the whole structure.

Finally, focusing on several, interacting structural features instead of statically attending to singular structural features is in accordance with the concept of resonance, requiring a global, dynamic view on the static structural formulas with keeping different structural features in mind in order to construct valid resonance structures (Nakhleh, 1992). Therefore, a predominant focus on singular (salient) structural features often leads to drawing mistakes (Cooper et al., 2010). Although familiarity and geometry may explain differences in the gaze distribution, fixations on structural features alone do not suffice to characterize productive and unproductive approaches when dealing with spatial complex molecules. For this reason, students' explanations of their drawing approach need to be taken into account.

RQ 3: How does a student's approach to visual selection to construct resonance structures relate to the productivity of the drawings?

Different visual selection approaches for the construction of resonance structures could be identified through students' verbal explanations for their drawings, which differ regarding the information selection approach (knowledge-driven or structure-driven) and the flexibility of these approaches (centered or variable). Thus, four categories can be distinguished: structure-driven centered, structure-driven variable, knowledge-driven centered, and knowledge-driven variable (**Figure 5**). **Figure 9** provides an overview of the absolute number of these approaches used for structure construction and the success rate for constructing a productive drawing. It can be derived from this table that about half of the drawings were constructed via a knowledge-driven variable approach, while only four drawings result from a structure-driven centered approach. Moreover, it can be seen that, contrary to our initial hypothesis, not a structure-driven or knowledge-driven approach, i.e., the explicit application of conceptual knowledge (Graulich et al., 2019), decides upon the productivity of the resulting drawing, but rather the

Visual selection approach	Absolute number	Productive (in %)	Unproductive (in %)
Structure-driven centered	4	25	75
Knowledge-driven centered	11	27.3	72.7
Structure-driven variable	10	90	10
Knowledge-driven variable	35	80	20

FIGURE 9
Absolute distribution and success rate of the different visual selection approaches.

flexibility of this approach, i.e., whether students focus on single structural features or take multiple structural features into account. At least 80% of the drawings with a variable approach are productive, whereas the amount of productive drawings reduces to under 30% if a centered approach has been used. In accordance with the results in the previous section, this shows that, besides being able to use resonance-related knowledge in the construction process (R-C factor), the success of constructing resonance structures heavily relies on students' overall ability to decode the structural formulas (R-M factor) (Schönborn and Anderson, 2008, 2010). How these four approaches influence the construction of resonance structures and how they are related to the drawing moves, can be illustrated by taking a closer, exemplary look at descriptions of students' drawing processes.

Nina, for instance, generated all her drawings for reaction A and B following a knowledge-driven centered approach (**Figure 10**), eventually resulting in unproductive structures, as the oxygen atom does not fulfill the octet rule, missing a positive charge on the oxygen atom in both cases. While expressing cluelessness about how to construct the resonance structures at first, Nina describes both her drawings by referring to experiences of delocalizing electrons of the oxygen atom as done in previous tasks. Therefore, the oxygen atom serves as an anchor guiding the drawing process and leading to multiple drawing moves, each starting from the oxygen atom and initiating the subsequent delocalization of the π-electrons of the double bond. As for the resonance structure in reaction A, she consequently misses a structure, which is, however, crucial for answering the task. Interestingly, while the initial positive charge remains unchanged in both structures, she creates a new, negative charge in each structure. The anchoring function of the oxygen atom at the extent of neglecting the other, more relevant structural features of the positive charge and double bond is also reflected in her attention distribution concerning the different structural features of the starting molecules in reaction A and B. **Figure 10** shows that Nina spends about 70% of the fixation time on the methoxy group in reaction A. In reaction B, she almost exclusively (95%) focuses on the methoxy group, not paying attention to the double bond.

A similar knowledge-driven centered approach can be seen in Catherine's drawing process of the resonance structure in reaction A as she centers the drawing process description solely on the function of the oxygen atom (**Figure 11**). As Nina, Catherine explains her drawing approach by emphasizing that the oxygen atom must participate in generating resonance structures due to the lone pairs that could be delocalized. This utterance shows that Catherine employs an overgeneralized rule (McClary and Talanquer, 2011) guiding her drawing process and making the oxygen atom her starting point for the construction. Therefore, her prior knowledge, i.e., the R-C factor, mainly drives the interpretation of the structural formula and the constructing of the resulting resonance structure (Schönborn and Anderson, 2008). Similar to Nina, this results in multiple drawing moves involving the delocalization of electrons of the oxygen atom across the double bond to form another double bond. While the positive charge remains unchanged at the carbon atom, the oxygen atom does not carry a charge, thus, resulting in an unproductive drawing. However, Catherine's fixation duration rate does not clearly reflect her drawing strategy (**Figure 11**). Despite spending much time on the methoxy group, Catherine also takes the double bond

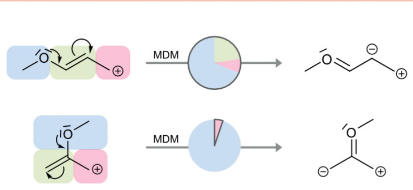


FIGURE 10
Nina's drawing moves, fixation duration rates on the structural features and drawing description for her drawings in reaction A and B. The orange frame indicates the unproductivity of the resulting drawings.

FIGURE 11
Catherine's drawing move, fixation duration rate on the structural features and drawing description for her drawing in reaction A. The orange frame indicates the unproductivity of the resulting drawing.

and positive charge into account, showing that she somewhat considered the whole molecule for the construction process.

In contrast to these examples, Elizabeth has a structure-driven centered approach to the construction of the resonance structure in reaction B (**Figure 12**). Similar to the other students, she centers her drawing description only on the double bond as single structural feature after having tried to form a carbon–oxygen π-bond. Elizabeth's approach illustrates a trial-and-error strategy (Ahmad and Omar, 1992), as she decides to delocalize the π-electrons of the double bond to the oxygen atom after struggling to move the electrons in the inverse direction. Although it becomes clear that she considers both structural features, as reflected in the fixation duration rates (**Figure 12**), she does not consider them in an interrelated manner but rather focuses on these structural features successively as singular entities. In Elizabeth's case, this approach results in a single drawing move through which additional charges are generated, and the oxygen atoms clearly breaks the octet rule. Interestingly, neither in her description nor in the fixation duration rate, the positive charge gets much attention. Therefore, she does not see the positive charge as a prerequisite for resonance in this task, but rather approaches such drawing tasks by considering electron-rich features, as she stated: *I guess lone pairs are also an indication for me that you can apply resonance, not only double bonds, but also lone pairs.*

So far, the centered approach shows that students justify their drawing processes by only referring to single structural features at the extent of neglecting other features within a molecule and considering single structural features as an anchor leading the entire drawing process. Many studies in STEM education research support our finding, that this anchoring function may be due to overgeneralized rules and heuristics, eventually causing algorithmic or arbitrary unproductive drawing steps and hindering the analysis of the given structure as a whole. In chemistry education, it has been shown that students are prone to focus on familiar and surface features, which eventually negatively impacts their problem solving (Kozma and Russell, 1997; Kraft et al., 2010; Anzovino and Bretz, 2015; Graulich and Bhattacharyya, 2017; Graulich et al., 2019) and supports their reliance on heuristics (McClary and Talanquer, 2011; Weinrich and Talanquer, 2015). Similarly, Inglis and Alcock (2012) have shown in the context of reading mathematical proofs that compared to mathematicians, undergraduate students spend more time focusing on surface features of an argument, i.e., they attend less to its logical structure.

In addition, these results may complement existing findings concerning the construction of structural formulas. As Ahmad and Omar (1992) and Cooper et al. (2010) stated, students often exhibit a trial-and-error approach or rely on memorized cues in drawing, which may result from the overreliance on singular structural features. Such a structural overreliance in the interpretation of structural representations (Schönborn and Anderson, 2008) may also offer a possible reason for the observed drawing difficulties of resonance structures reported by Betancourt-Pérez et al. (2010) and Petterson et al. (2020).

In contrast to the centered approach, the variable approach is characterized by a more holistic approach to the drawing process of different resonance structures. Students showing this approach take multiple, interrelated drawing features into account. For instance, consider Paula's approach in reaction A (**Figure 13**).

Paula exhibits a structure-driven approach by merely referring to interacting structural features, i.e., the positive charge in both resonance structures and the electron-rich double bond or the lone pairs of the oxygen atom. However, in

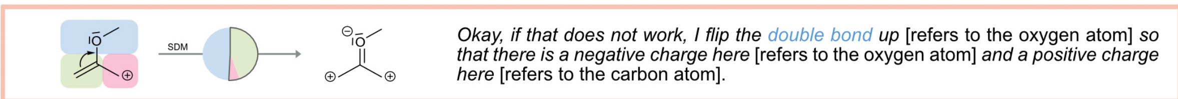


FIGURE 12
Elizabeth's drawing move, fixation duration rate on the structural features and drawing description for her drawing in reaction B. The orange frame indicates the unproductivity of the resulting drawing.

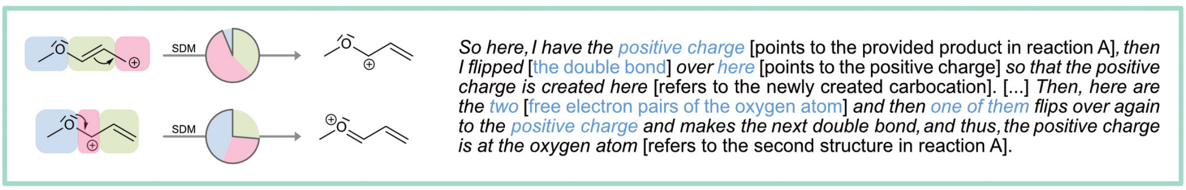


FIGURE 13
Paula's drawing moves, fixation duration rates on the structural features and drawing description for her drawings in reaction A. The green frame indicates the productivity of the resulting drawings.

10.3389/feduc.2022.1055280

contrast to Elizabeth's approach, she does not just focus on one feature, but sequentially considers smaller, interacting parts of the molecule which may aid in the delocalization of the positive charge. Therefore, as **Figure 13** illustrates, this approach leads to single drawing moves delocalizing electron density toward an electron-deficient atom. This sequential approach is also reflected in Paula's fixation duration rate, as she focuses most of the time on the positive charge and the double bond in the first resonance structure, and then, on the newly created positive charge and the methoxy group in the second structure.

Figure 9 illustrates that 35 of 60 drawings stem from a knowledge-driven variable approach, in which rules and concepts, such as stability or electronic effects, guided students' drawing process. Phil's sequential drawing process in reaction A and B exemplifies this approach. As **Figure 14** depicts, each productive resonance structure results from a single drawing move, which shows the movement of electrons from an electron-rich source to an electron-deficient atom. In his verbalization of the drawing process, Phil repeatedly refers to the delocalization of the positive charge and stability as the driving force for the generation of each structure.

It can be inferred from Phil's drawing approach that he centers his drawing moves around the positive charge. He takes a rather analytic, holistic approach by considering adjacent structural features and analyzing their contribution to the delocalization of the positive charge. Consider therefore the fixation duration rate during the construction process of the different resonance structures, exhibiting an attention distribution trend for the structures in reaction A (**Figure 14**). While Phil mainly fixates the double bond and the positive charge in the first resonance structure, his attention is drawn to the methoxy group and the positive charge in the second structure. In contrast to that, he centers his attention on the methoxy group and the double bond in the resonance structure in reaction B. Given that he verbalizes stability as the driving force for the construction of the structures, this gaze distribution may be a result of the analysis and weighing of the structural features and their influences on resonance. Therefore, Phil's knowledge-driven variable approach shows that he reflects upon his drawing moves. This illustrates the need for the successful intertwined application of the different factors (R, C, M) in order to interpret a structural representation and subsequently construct another resonance structure (Schönborn and Anderson, 2008). This reflective drawing approach becomes even more apparent for Luke, another student who has a knowledge-driven variable approach. Concerning a resonance structure in A, he describes in detail the electronic effect of the oxygen atom and thereafter weighs the overall stability of the resulting resonance structure:

> *Then I thought, what does the oxygen atom do? The oxygen atom has a negative inductive effect, which is why the positive charge is intensified. But I thought that it can be neglected given the positive resonance effect, because the electron pair can be pushed toward the positive charge, and then, the positive charge would be here. I would say that is energetically not so favorable here with the positive charge on the oxygen, but it shows that the positive charge can be distributed relatively well.*

Given that the gaze proportion of productive and unproductive drawings does not differ much in reaction B (cf. section "RQ 2: What structural features do students attend to in terms of attention distribution when translating one resonance structure into another and how is it related to students' drawing moves?"), it may be inferred from these examples that students process the structural information differently by applying their related knowledge in different ways. While students with an unproductive drawing may tend to look at the given structure statically and apply thereon rules, for productive drawings, it may be the case that a higher gaze proportion indicates a more thorough weighing and

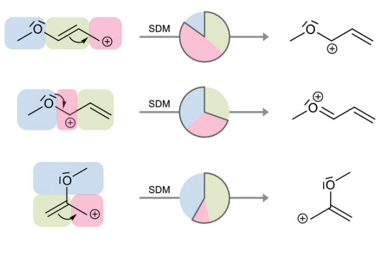


FIGURE 14
Phil's drawing moves, fixation duration rates on the structural features and drawing description for his drawings in reaction A and B. The green frame indicates the productivity of the resulting drawings.

reflection of its contribution to the drawing move. Hence, this inference strengthens Keig and Rubba's (1993) finding that solving information-processing tasks (e.g., translations between structures) requires a thorough understanding (and thus application) of the underlying concept.

Altogether, it can be derived from these examples that students with a variable approach demonstrate more flexibility when constructing resonance structures by showing a more sequential, distributed attention to structural features, often resulting in productive single drawing moves. By being analytic in nature, they exhibit a more expert-like task approach (Stieff and Raje, 2008). By successively analyzing structural features and their relationship in terms of constructing a subsequent resonance structure, students consequently demonstrate both a local and global coherence formation (Seufert, 2003), altogether resulting in a holistic approach.

Conclusion and implications

This study is the first to explore in-depth students' drawing processes in organic chemistry with the help of eye-tracking. It aimed at providing process-oriented insights into how students connect drawing-related information during the construction of resonance structures given students' struggle with these representations. To this end, we analyzed in detail students' transition patterns, the co-occurrence of fixated drawing elements during the construction process, students' gaze proportion on structural features, and shed light on their visual selection approaches to navigating the drawing processes.

As the main results, we found that on a global level, students exhibit a similar gaze behavior concerning information retrieval and integration, irrespective of the productivity of their drawing, i.e., the same amount of information was used for the construction of resonance structures. However, the transition types distinguished a productive drawing from an unproductive drawing. While productive drawings result from more transitions between previous and target drawings, the construction of unproductive drawings is characterized by more transitions between target and unrelated drawings (RQ 1). Due to a predominant connection of previous and target drawings across all resonance structures, the analysis of the gaze proportion on structural features of the previous drawing revealed that productive structures are characterized by a higher sequential fixation on interrelated structural features. In contrast, a tendency of focusing single (unrelated) structural features emerged for unproductive drawings. However, this difference did not apply to resonance structures that were spatially more complex (i.e., a non-linear, branched molecule) (RQ 2). Finally, a qualitative look at students' visual selection approaches, as reflected by students' descriptions of drawing processes, showed that a variable approach underlies many productive drawings, i.e., an analytical approach, in which students attended to interrelated, relevant structural features. This is contrary to a centered approach, resulting in more unproductive drawings; here, students focused on singular structural features in a static manner (RQ 3). These findings bear different implications for both instruction and research.

Given that students with unproductive drawings often connect and focus on unrelated drawing elements, when, however, a holistic, analytical approach is required, instructional interventions should aim more at directing students' attention, thus, supporting them to assess given information such as structural features, their relevance and their role within the construction of resonance structures (e.g., as possible starting points), therefore foster students' ability to decode given representations (R-M factor) (Schönborn and Anderson, 2008, 2010). In this regard, possible instructional interventions could use process-oriented highlighting of interacting structural features, e.g., via tutorial videos (e.g., Rodemer et al., 2021). In this way, the learners' attention could be better directed, as they can see and follow the actual drawing process instead of only seeing the static resonance structures as the final result on paper. As a possible intervention to externalize the viewing process, eye-movement-modeling examples could be used as worked examples (van Gog et al., 2009; Jarodzka et al., 2013). In addition, the results indicate that students' conceptual knowledge of the resonance concept, specifically its flexible application, seems to be of great importance with respect to drawing productive resonance structures, thus, besides the R-M factor, emphasis should also be placed on the R-C factor (Schönborn and Anderson, 2008). Consequently, teaching algorithms *how* to draw such structures is not enough for building sustainable drawing skills. Instead, more effort should be put into addressing *why* certain drawing steps occur, e.g., what chemical concepts the drawing steps build upon. Creating such interventions would make students more reflective in assessing drawing moves. In turn, this could help students' ability to flexibly decide when resonance applies and reduce their overreliance on heuristics and rules, such as searching for familiar or salient surface features.

As this study is explorative, more research in other contexts, with different task designs, and with more participants is needed to complement our findings and test whether our results can be confirmed. By finding little differences in eye movements while dealing with the more complex structure in reaction B, the construction of more complex structures deserves more attention. Due to their complexity and relevance, aromatic compounds are suitable for this purpose. However, more complex structures complicate the comparative analysis because, depending on the complexity, the number of differing drawing products increases. At the same time, the use of simpler aromatic compounds may lead students to proceed algorithmically. For instance, to investigate the differences in the information retrieval of drawing-related structural features, multiple choice tasks could be used, in which learners are asked to decide on a resulting structure based on an initial compound. As such, the comparative investigation of differing fixation distributions when choosing a productive or unproductive structure would be possible. Likewise, in such a setting, a stationary eye-tracker could be used, which provides more precise results regarding the fixation of single bonds and structural features.

By inferring underlying processes for previously documented drawing difficulties, this study adds to existing research by focusing on the context of the resonance concept. Furthermore, given its process-oriented character, this study adds to existing STEM-related drawing research in general. Despite previous research on the effects of drawing activities and the drawing products' quality on learning outcomes, few research focused on the construction process underlying learners' drawing products so far (Lobato et al., 2014). However, a process-oriented perspective can provide additional diagnostic insights into representation-related difficulties as the drawing processes can reveal learners' scientific thoughts and conceptions (Lobato et al., 2014) as well as their unconscious actions when generating representations. As this study exemplified, the use of eye-tracking showed that, besides conceptual knowledge, students' ability to construct productive scientific representations crucially depends on their competence to decode and manipulate such representations. As such, this methodology could be also applied in other contexts and STEM disciplines to reveal cognitive processes underlying the generation of representations, e.g., what features learners (unconsciously) use and integrate to depict submicroscopic processes in biology or to construct diagrams in physics, and whether the use of specific features or drawing elements influences the overall drawing quality. Eye-tracking could also be used in other contexts in organic chemistry, such as observing learners' stepwise construction of mechanisms and the use of the electron-pushing formalism (e.g., Bhattacharyya and Bodner, 2005; Grove et al., 2012; Flynn and Featherstone, 2017), or observing the order and linearity of the diagram construction depending on the learners' expertise. Although our work yields first process-oriented insights, further research across different scientific disciplines is needed to provide a comprehensive picture of students' drawing approaches and difficulties for scientific representations.

Limitations

This study was meant to be exploratory in nature and aimed at offering insights into students' drawing processes in organic chemistry. Therefore, some limitations with respect to the analysis and results must be considered. First, the small number of students ($N = 20$) and the focus on one task might limit the generalizability of the findings. Additionally, students were required to construct only two additional resonance structures in the first, and one structure in the second reaction, both dealing with rather simple structural formulas. Unlike in the first reaction, the construction of resonance structures did not build upon each other in the second reaction. Given these constraints, it remains unclear to which extent our findings apply to drawing processes of more complex structures. Furthermore, case comparisons might have motivated students to include and compare information of both reactions in their drawing process, since the given structures only differ in their connectivity. Thus, this variable may also have influenced students' drawing processes, specifically on the integration of unrelated information. Concerning the interviews, the participants were prompted to describe their drawing process. However, it is possible that they used knowledge that they did not verbalize and which, thus, remained implicit.

Finally, technical limitations with respect to the mobile eye-tracker must be considered. The eye-tracker exhibits technical measurement inaccuracies due to, among others, a variable field of view, movement of the head, and the slippage of the glasses as a result of a longer measurement period. These inaccuracies can affect the results and the data analysis (e.g., requiring the correction of systematic gaze point offset by adjusting the AOIs).

Data availability statement

The datasets presented in this article cannot be made openly available because of the privacy policy stated in the participants' consent form. Requests to access the datasets should be directed to the corresponding author NG, Nicole.Graulich@didaktik.chemie.uni-giessen.de.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

IB designed the study with the help of NG, collected the data, and wrote the manuscript. IB and AL analyzed the data. IB, AL, and NG reviewed, edited, and discussed the manuscript and the analyses. All authors contributed to the article and approved the submitted version.

Funding

This work was funded in part by the National Science Foundation (DRL-2100320, DRL-2201723, and DRL-2225240), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

Acknowledgments

This publication represents a component of the first author's doctoral (Dr. rer. nat.) thesis in the Faculty of Biology and Chemistry at the Justus Liebig University Giessen, Germany. We thank all students who participated in the study as well as the members of the Graulich Research Group for fruitful discussions about the research. We are grateful to Ramon Buchczyk and Helena Rist for support in data analysis. In the study, we utilized the Epistemic network analysis (ENA) method.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, W.-Y., and Omar, S. (1992). Drawing Lewis structures: A step-by-step approach. *J. Chem. Educ.* 69, 791–792. doi: 10.1021/ed069p791

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learn. Instr.* 16, 183–198. doi: 10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Prain, V., and Tytler, R. (2011). Drawing to learn in science. *Science* 333, 1096–1097. doi: 10.1126/science.1204153

Ainsworth, S. E., and Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Curr. Dir. Psychol. Sci.* 30, 61–67. doi: 10.1177/0963721420979582

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *J. Vis.* 13, 1–16. doi: 10.1167/13.3.7

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., and Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Front. Psychol.* 6:1016. doi: 10.3389/fpsyg.2015.01016

Anzovino, M. E., and Bretz, S. L. (2015). Organic chemistry students' ideas about nucleophiles and electrophiles: The role of charges and mechanisms. *Chem. Educ. Res. Pract.* 16, 797–810. doi: 10.1039/C5RP00113G

Awh, E., Belopolsky, A. V., and Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends Cogn. Sci.* 16, 437–443. doi: 10.1016/j.tics.2012.06.010

Baluyut, J. Y., and Holme, T. A. (2019). Eye tracking student strategies for solving stoichiometry problems involving particulate nature of matter diagrams. *Chem. Teach. Int.* 1, 1–13. doi: 10.1515/cti-2018-0003

Betancourt-Pérez, R., Olivera, L. J., and Rodríguez, J. E. (2010). Assessment of organic chemistry students' knowledge of resonance-related structures. *J. Chem. Educ.* 87, 547–551. doi: 10.1021/ed800163g

Bhattacharyya, G., and Bodner, G. M. (2005). "It gets me to the product": How students propose organic mechanisms. *J. Chem. Educ.* 82, 1402–1407. doi: 10.1021/ed082p1402

Bobek, E., and Tversky, B. (2016). Creating visual explanations improves learning. *Cogn. Res. Princ. Implic.* 1:27. doi: 10.1186/s41235-016-0031-6

Bodner, G. M., and Domin, D. S. (2000). Mental models: The role of representations in problem solving in chemistry. *Univ. Chem. Educ.* 4, 24–30.

Brandfonbrener, P. B., Watts, F. M., and Shultz, G. V. (2021). Organic chemistry students' written descriptions and explanations of resonance and its influence on reactivity. *J. Chem. Educ.* 98, 3431–3441. doi: 10.1021/acs.jchemed.1c00660

Brennan, R. L., and Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educ. Psychol. Meas.* 41, 687–699. doi: 10.1177/001316448104100307

Bruckner, S., Schneider, J., Zlatkin-Troitschanskaia, O., and Drachsler, H. (2020). Epistemic network analyses of economics students' graph understanding: An eye-tracking study. *Sensors* 20, 1–33. doi: 10.3390/s20236908

Carle, M. S., and Flynn, A. B. (2020). Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chem. Educ. Res. Pract.* 21, 622–637. doi: 10.1039/C9RP00203K

Cartrette, D. P., and Bodner, G. M. (2009). Non-mathematical problem solving in organic chemistry. *J. Res. Sci. Teach.* 47, 643–660. doi: 10.1002/tea.20306

Connor, M. C., Glass, B. H., Finkenstaedt-Quinn, S. A., and Shultz, G. V. (2021). Developing expertise in (1)H NMR spectral interpretation. *J. Org. Chem.* 86, 1385–1395. doi: 10.1021/acs.joc.0c01398

Cooper, M. M., Grove, N., Underwood, S. M., and Klymkowsky, M. W. (2010). Lost in Lewis structures: An investigation of student difficulties in developing representational competence. *J. Chem. Educ.* 87, 869–874. doi: 10.1021/ed900004y

Cooper, M. M., Stieff, M., and DeSutter, D. (2017). Sketching the invisible to predict the visible: From drawing to modeling in chemistry. *Top. Cogn. Sci.* 9, 902–920. doi: 10.1111/tops.12285

Cooper, M. M., Underwood, S. M., Hilley, C. Z., and Klymkowsky, M. W. (2012). Development and assessment of a molecular structure and properties learning progression. *J. Chem. Educ.* 89, 1351–1357. doi: 10.1021/ed300083a

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learn. Instr.* 9, 343–363. doi: 10.1016/S0959-4752(98)00051-6

Cromley, J. G., Du, Y., and Dane, A. P. (2019). Drawing-to-learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing? *J. Sci. Educ. Technol.* 29, 216–229. doi: 10.1007/s10956-019-09807-6

Cullipher, S., Hansen, S. J., and VandenPlas, J. R. (2018). "Eye tracking as a research tool: An introduction," in *Eye tracking for the chemistry education researcher*, eds J. R. VandenPlas, S. J. R. Hansen, and S. Cullipher (Washington, DC: ACS Publications), 1–9. doi: 10.1021/bk-2018-1292.ch001

Cullipher, S., and VandenPlas, J. R. (2018). "Using fixations to measure attention," in *Eye tracking for the chemistry education researcher*, eds J. R. VandenPlas, S. J. R. Hansen, and S. Cullipher (Washington, DC: ACS Publications), 53–72. doi: 10.1021/bk-2018-1292.ch004

DeFever, R. S., Bruce, H., and Bhattacharyya, G. (2015). Mental rolodexing: Senior chemistry majors' understanding of chemical and physical properties. *J. Chem. Educ.* 92, 415–426. doi: 10.1021/ed500360g

Dood, A. J., and Watts, F. M. (2022). Mechanistic reasoning in organic chemistry: A scoping review of how students describe and explain mechanisms in the chemistry education research literature. *J. Chem. Educ.* 99, 2864–2876. doi: 10.1021/acs.jchemed.2c00313

Duis, J. M. (2011). Organic chemistry educators' perspectives on fundamental concepts and misconceptions: An exploratory study. *J. Chem. Educ.* 88, 346–350. doi: 10.1021/ed1007266

Fiorella, L., Stull, A. T., Kuhlmann, S., and Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *J. Educ. Psychol.* 112, 895–906. doi: 10.1037/edu0000408

Fiorella, L., and Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educ. Psychol. Rev.* 30, 1115–1137. doi: 10.1007/s10648-018-9444-8

Flynn, A. B., and Featherstone, R. B. (2017). Language of mechanisms: Exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions. *Chem. Educ. Res. Pract.* 18, 64–77. doi: 10.1039/c6rp00126b

Gegenfurtner, A., Lehtinen, E., and Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educ. Psychol. Rev.* 23, 523–552. doi: 10.1007/s10648-011-9174-7

Gilbert, J. K. (2005). "Visualization: A metacognitive skill in science and science education," in *Visualization in Science Education*, ed. J. K. Gilbert (New York, NY: Springer), 9–27.

Gobert, J. D., and Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *J. Res. Sci. Teach.* 36, 39–53.

Goodwin, W. M. (2008). Structural formulas and explanation in organic chemistry. *Found. Chem.* 10, 117–127. doi: 10.1007/s10698-007-9033-2

Graulich, N. (2015). The tip of the iceberg in organic chemistry classes: How do students deal with the invisible? *Chem. Educ. Res. Pract.* 16, 9–21. doi: 10.1039/C4RP00165F

Graulich, N., and Bhattacharyya, G. (2017). Investigating students' similarity judgments in organic chemistry. *Chem. Educ. Res. Pract.* 18, 774–784. doi: 10.1039/C7RP00055C

Graulich, N., Hedtrich, S., and Harzenetter, R. (2019). Explicit versus implicit similarity–exploring relational conceptual understanding in organic chemistry. *Chem. Educ. Res. Pract.* 20, 924–936. doi: 10.1039/C9RP00054B

Graulich, N., and Schween, M. (2018). Concept-oriented task design: Making purposeful case comparisons in organic chemistry. *J. Chem. Educ.* 95, 376–383. doi: 10.1021/acs.jchemed.7b00672

Grove, N. P., Cooper, M. M., and Rush, K. M. (2012). Decorating with arrows: Toward the development of representational competence in organic chemistry. *J. Chem. Educ.* 89, 844–849. doi: 10.1021/ed2003934

Havelková, L., and Gołębiowska, I. M. (2019). What went wrong for bad solvers during thematic map analysis? Lessons learned from an eye-tracking study. *ISPRS Int. J. Geo Inf.* 9:9. doi: 10.3390/ijgi9010009

Hejnová, E., and Kekule, M. (2018). Observing students' problem solving strategies in mechanics by eye-tracking method. *Sci. Educ.* 9, 102–116. doi: 10.14712/18047106.1018

Hellenbrand, J., Mayer, R. E., Opfermann, M., Schmeck, A., and Leutner, D. (2019). How generative drawing affects the learning process: An eye-tracking analysis. *Appl. Cogn. Psychol.* 33, 1147–1164. doi: 10.1002/acp.3559

Holmqvist, K., Andrà, C., Lindström, P., Arzarello, F., Ferrara, F., Robutti, O., et al. (2011). A method for quantifying focused versus overview behavior in AOI sequences. *Behav. Res. Methods* 43, 987–998. doi: 10.3758/s13428-011-0104-x

Inglis, M., and Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *J. Res. Math. Educ.* 43, 358–390. doi: 10.5951/jresematheduc.43.4.0358

Jarodzka, H., Scheiter, K., Gerjets, P., and van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learn. Instr.* 20, 146–154. doi: 10.1016/j.learninstruc.2009.02.019

Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., and Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learn. Instr.* 25, 62–70. doi: 10.1016/j.learninstruc.2012.11.004

Johnson, C. I., and Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *J. Exp. Psychol. Appl.* 18, 178–191. doi: 10.1037/a0026923

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *J. Comput. Assist. Learn.* 7, 75–83. doi: 10.1111/j.1365-2729.1991.tb00230.x

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037/0033-295X.87.4.329

Karonen, M., Murtonen, M., Södervik, I., Manninen, M., and Salomäki, M. (2021). Heuristics hindering the development of understanding of molecular structures in university level chemistry education: The Lewis structure as an example. *Educ. Sci.* 11:258. doi: 10.3390/educsci11060258

Kaufmann, I., Hamza, K. M., Rundgren, C.-J., and Eriksson, L. (2017). Developing an approach for teaching and learning about Lewis structures. *Int. J. Sci. Educ.* 39, 1601–1624. doi: 10.1080/09500693.2017.1342053

Keig, P. F., and Rubba, P. A. (1993). Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *J. Res. Sci. Teach.* 30, 883–903. doi: 10.1002/tea.3660300807

Kim, T., Wright, L. K., and Miller, K. (2019). An examination of students' perceptions of the Kekulé resonance representation using a perceptual learning theory lens. *Chem. Educ. Res. Pract.* 20, 659–666. doi: 10.1039/C9RP00009G

Kozma, R., Chin, E., Russell, J., and Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *J. Learn. Sci.* 9, 105–143. doi: 10.1207/s15327809jls0902_1

Kozma, R., and Russell, J. (2005). "Students becoming chemists: Developing representational competence," in *Visualization in Science Education*, ed. J. K. Gilbert (New York, NY: Springer), 121–145.

Kozma, R. B., and Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *J. Res. Sci. Teach.* 34, 949–968.

Kraft, A., Strickland, A. M., and Bhattacharyya, G. (2010). Reasonable reasoning: Multi-variate problem-solving in organic chemistry. *Chem. Educ. Res. Pract.* 11, 281–292. doi: 10.1039/C0RP90003F

Kuckartz, U., and Rädiker, S. (2019). *Analyzing Qualitative Data with MAXQDA*. New York, NY: Springer.

Leopold, C., and Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learn. Instr.* 22, 16–26. doi: 10.1016/j.learninstruc.2011.05.005

Lesh, R., Post, T. R., and Behr, M. (1987). "Representations and translations among representations in mathematics learning and problem solving," in *Problems of Representations in the Teaching and Learning of Mathematics*, ed. C. Janvier (Hillsdale, NJ: Lawrence Erlbaum), 33–40.

Lobato, J., Hohensee, C., and Diamond, J. M. (2014). What can we learn by comparing students' diagram-construction processes with the mathematical conceptions inferred from their explanations with completed diagrams? *Math. Educ. Res. J.* 26, 607–634. doi: 10.1007/s13394-013-0106-3

Malone, S., Altmeyer, K., Vogel, M., and Brünken, R. (2020). Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study. *J. Comput. Assist. Learn.* 36, 781–798. doi: 10.1111/jcal.12426

Maries, A., and Singh, C. (2018). Do students benefit from drawing productive diagrams themselves while solving introductory physics problems? The case of two electrostatics problems. *Eur. J. Phys.* 39:015703. doi: 10.1088/1361-6404/aa9038

Mason, L., Lowe, R., and Tornatora, M. C. (2013). Self-generated drawings for supporting comprehension of a complex animation. *Contemp. Educ. Psychol.* 38, 211–224. doi: 10.1016/j.cedpsych.2013.04.001

Mayer, R. E. (2005). "Cognitive theory of multimedia learning," in *The Cambridge Handbook of Multimedia Learning*, ed. R. E. Mayer (Cambridge, MA: Cambridge University Press), 31–48.

McClary, L., and Talanquer, V. (2011). Heuristic reasoning in chemistry: Making decisions about acid strength. *Int. J. Sci. Educ.* 33, 1433–1454. doi: 10.1080/09500693.2010.528463

Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *J. Chem. Educ.* 69, 191–196. doi: 10.1021/ed069p191

Nitz, S., Ainsworth, S. E., Nerdel, C., and Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learn. Instr.* 31, 13–22. doi: 10.1016/j.learninstruc.2013.12.003

Olimpo, J. T., Kumi, B. C., Wroblewski, R., and Dixon, B. L. (2015). Examining the relationship between 2D diagrammatic conventions and students' success on representational translation tasks in organic chemistry. *Chem. Educ. Res. Pract.* 16, 143–153. doi: 10.1039/C4RP00169A

Ott, N., Brünken, R., Vogel, M., and Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learn. Instr.* 58, 88–105. doi: 10.1016/j.learninstruc.2018.04.010

Petterson, M. N., Watts, F. M., Snyder-White, E. P., Archer, S. R., Shultz, G. V., and Finkenstaedt-Quinn, S. A. (2020). Eliciting student thinking about acid–base reactions via app and paper–pencil based problem solving. *Chem. Educ. Res. Pract.* 21, 878–892. doi: 10.1039/C9RP00260J

Quillin, K., and Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE Life Sci. Educ.* 14:es2. doi: 10.1187/cbe.14-08-0128

Rellensmann, J., Schukajlow, S., and Leopold, C. (2016). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educ. Stud. Math.* 95, 53–78. doi: 10.1007/s10649-016-9736-1

Richardson, W. (1986). Teaching the concept of resonance with transparent overlays. *J. Chem. Educ.* 63, 518–519. doi: 10.1021/ed063p518

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2020). Decoding case comparisons in organic chemistry: Eye-tracking students' visual behavior. *J. Chem. Educ.* 97, 3530–3539. doi: 10.1021/acs.jchemed.0c00418

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2021). Connecting explanations to representations: Benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry. *Int. J. Sci. Educ.* 43, 2707–2728. doi: 10.1080/09500693.2021.1985743

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Research*. Newbury Park, CA: Sage Publications.

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers*. Los Angeles: Sage Publications.

Sandi-Urena, S., Cambronero, G. L., and Chaves, D. J. (2020). Conceptualisation of Lewis structures by chemistry majors. *Chem. Teach. Int.* 2, 1–9. doi: 10.1515/cti-2018-0019

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., and Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemp. Educ. Psychol.* 39, 275–286. doi: 10.1016/j.cedpsych.2014.07.003

Schmidgall, S. P., Scheiter, K., and Eitel, A. (2020). Can we further improve tablet-based drawing to enhance learning? An empirical test of two types of support. *Instr. Sci.* 48, 453–474. doi: 10.1007/s11251-020-09513-6

Schmidt-Weigand, F., Kohnert, A., and Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learn. Instr.* 20, 100–110. doi: 10.1016/j.learninstruc.2009.02.011

Schönborn, K. J., and Anderson, T. R. (2008). A model of factors determining students' ability to interpret external representations in biochemistry. *Int. J. Sci. Educ.* 31, 193–232. doi: 10.1080/09500690701670535

Schönborn, K. J., and Anderson, T. R. (2010). Bridging the educational research-teaching practice gap: Foundations for assessing and developing biochemistry students' visual literacy. *Biochem. Mol. Biol. Educ.* 38, 347–354. doi: 10.1002/bmb.20436

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., and Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *J. Educ. Psychol.* 102, 872–879. doi: 10.1037/a0019640

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learn. Instr.* 13, 227–237. doi: 10.1016/S0959-4752(02)00022-1

Seufert, T., and Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Appl. Cogn. Psychol.* 20, 321–331. doi: 10.1002/acp.1248

Shaffer, D. W., Collier, W., and Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *J. Learn. Anal.* 3, 9–45. doi: 10.18608/jla.2016.33.3

Shaffer, D. W., and Ruis, A. R. (2017). "Epistemic network analysis: A worked example of theory-based learning analytics," in *Handbook of Learning Analytics*, eds C. Lang, G. Siemens, A. Wise, and D. Gašević (New York, NY: Society for Learning Analytics Research), 175–187. doi: 10.18608/hla17.015

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. doi: 10.2307/2333709

Stieff, M., and DeSutter, D. (2020). Sketching, not representational competence, predicts improved science learning. *J. Res. Sci. Teach.* 58, 128–156. doi: 10.1002/tea.21650

Stieff, M., and Raje, S. (2008). "Expertise and spatial reasoning in advanced scientific problem solving," in *International Perspectives in the Learning Sciences: Creating a learning world. Proceedings of the Eighth International Conference for the Learning Sciences ICLS*, eds G. Kanselaar, V. Jonker, P. A. Kirschner, and F. J. Prins (Utrecht: The Netherlands: International Society of the Learning Sciences), 366–373.

Strickland, A. M., Kraft, A., and Bhattacharyya, G. (2010). What happens when representations fail to represent? Graduate students' mental models of organic chemistry diagrams. *Chem. Educ. Res. Pract.* 11, 293–301. doi: 10.1039/C0RP90009E

Taber, K. S. (2002). Compounding quanta: Probing the frontiers of student understanding of molecular orbitals. *Chem. Educ. Res. Pract.* 3, 159–173. doi: 10.1039/B2RP90013K

Tang, H., Day, E. L., Atkinson, M. B., and Pienta, N. J. (2018). GrpString: An R package for analysis of groups of strings. *R J.* 10, 359–369. doi: 10.32614/RJ-2018-002

Tang, H., Kirk, J., and Pienta, N. J. (2014). Investigating the effect of complexity factors in stoichiometry problems using logistic regression and eye tracking. *J. Chem. Educ.* 91, 969–975. doi: 10.1021/ed4004113

Tang, H., and Pienta, N. (2012). Eye-tracking study of complexity in gas law problems. *J. Chem. Educ.* 89, 988–994. doi: 10.1021/ed200644k

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Curr. Opin. Psychol.* 29, 97–101. doi: 10.1016/j.copsyc.2018.12.024

Tiettmeyer, J. M., Coleman, A. F., Balok, R. S., Gampp, T. W., Duffy, P. L., Mazzarone, K. M., et al. (2017). Unraveling the complexities: An investigation of the factors that induce load in chemistry students constructing Lewis structures. *J. Chem. Educ.* 94, 282–288. doi: 10.1021/acs.jchemed.6b00363

Topczewski, J. J., Topczewski, A. M., Tang, H., Kendhammer, L. K., and Pienta, N. J. (2016). NMR spectra through the eyes of a student: Eye tracking applied to NMR items. *J. Chem. Educ.* 94, 29–37. doi: 10.1021/acs.jchemed.6b00528

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., and Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Comput. Hum. Behav.* 25, 785–791. doi: 10.1016/j.chb.2009.02.007

Van Meter, P., and Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educ. Psychol. Rev.* 17, 285–325. doi: 10.1007/s10648-005-8136-3

Weinrich, M., and Talanquer, V. (2015). Mapping students' conceptual modes when thinking about chemical reactions used to make a desired product. *Chem. Educ. Res. Pract.* 16, 561–577. doi: 10.1039/C5RP00024F

Wu, S. P., and Rau, M. A. (2019). How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities. *Educ. Psychol. Rev.* 31, 87–120. doi: 10.1007/s10648-019-09467-3

Xue, D., and Stains, M. (2020). Exploring students' understanding of resonance and its relationship to instruction. *J. Chem. Educ.* 97, 894–902. doi: 10.1021/acs.jchemed.0c00066

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 12 December 2022
DOI 10.3389/feduc.2022.1003740

Check for updates

OPEN ACCESS

EDITED BY
Dor Abrahamson,
University of California, Berkeley,
United States

REVIEWED BY
Alik Palatnik,
Hebrew University of Jerusalem, Israel
Rotem Abdu,
David Yellin College of Education,
Israel

*CORRESPONDENCE
Aylin Thomaneck
ay_th@uni-bremen.de

SPECIALTY SECTION
This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

RECEIVED 26 July 2022
ACCEPTED 21 November 2022
PUBLISHED 12 December 2022

CITATION
Thomaneck A, Vollstedt M and Schindler M (2022) What can eye movements tell about students' interpretations of contextual graphs? A methodological study on the use of the eye-mind hypothesis in the domain of functions. *Front. Educ.* 7:1003740. doi: 10.3389/feduc.2022.1003740

What can eye movements tell about students' interpretations of contextual graphs? A methodological study on the use of the eye-mind hypothesis in the domain of functions

Aylin Thomaneck[1]*, Maike Vollstedt[1] and Maike Schindler[2]

[1]Department of Mathematics and Computer Science, University of Bremen, Bremen, Germany,
[2]Department of Rehabilitation and Special Education, University of Cologne, Cologne, Germany

Introduction: The use of eye tracking (ET) in mathematics education research has increased in recent years. Eye tracking is a promising research tool in the domain of functions, especially in graph interpretation. It promises to gain insights into learners' approaches and ways of thinking. However, for the domain of functions and graph interpretation, it has not yet been investigated how eye-tracking data can be interpreted. In particular, it is not clear how eye movements may reflect students' cognitive processes. Thus, in this study, we investigate in how far the eye-mind hypothesis (EMH), which states broadly that what the eye fixates is currently being processed, can be applied to this subdomain. This is particularly true for contextual graphs, whose data originate from real-world situations, and which are of central importance for the development of mathematical literacy. The aim of our research is to investigate how eye movements can be interpreted in the domain of functions, particularly in students' interpretations of contextual graphs.

Methods: We conducted an exploratory case study with two university students: The students' eye movements were recorded while they worked on graph interpretation tasks in three situational contexts at different question levels. Additionally, we conducted subsequent stimulated recall interviews (SRIs), in which the students recalled and reported their original thoughts while interpreting the graphs.

Results: We found that the students' eye movements were often related to students' cognitive processes, even if indirectly at times, and there was only limited ambiguity in the interpretation of eye movements. However, we also found domain-specific as well as domain-general challenges in interpreting eye movements.

Discussion: Our results suggest that ET has a high potential to gain insights into students' graph interpretation processes. Furthermore, they point out what aspects, such as ambiguity and peripheral vision, need to be taken into consideration when investigating eye movements in the domain of functions.

KEYWORDS

eye tracking (ET), eye-mind hypothesis, eye movements, functions, graph interpretation, contextual graphs, stimulated recall

Introduction

The use and relevance of eye tracking (ET), the capturing of a person's eye movements using ET devices, in research significantly increased in recent years (König et al., 2016). In mathematics education, too, ET is gaining interest and is used in numerous areas and contexts (Strohmaier et al., 2020). Whereas some ET studies take an embodied perspective on eye movements, in that mind and eye movements are considered as parts of the body as an entity (e.g., Abrahamson and Bakker, 2016), other ET studies take a more psychological perspective by understanding eye movements "as a window to cognition" (König et al., 2016, p. 2). However, in both frameworks, ET is used in particular to investigate students' thinking and learning processes. Therefore, many studies rely on the eye-mind hypothesis (EMH) (Just and Carpenter, 1976b), which presumes a close relationship between what persons fixate on and what they process. However, studies from various fields revealed several limitations of this assumption (Underwood and Everatt, 1992; Anderson et al., 2004; Kliegl et al., 2006; Schindler and Lilienthal, 2019; Wu and Liu, 2022). Therefore, the relationship between eye movements and cognitive processes should be examined carefully for every subdomain to reduce the inherent ambiguity and uncertainty in interpreting gaze data.

Our study follows up on this uncertainty regarding the application of the EMH and the need to investigate domain-relatedly how eye movements can be interpreted. We focus on the relationship between eye movements and cognition in the domain of functions. In general, the domain of functions is an interesting field to study for the following reason: It is characterized by having very high relevance for both everyday life and school lessons (Friel et al., 2001). Especially in the first case, functions are often related to a situational context, for example, when they are used to model empirical phenomena. In our study, we are interested in the interpretation of graphs. Specifically, we study what we call *contextual graphs*. These are graphs whose data originate from measured values of real-world situations. In the digital age, graphs are pervasive in society and in everyday lives and, thus, also an important topic in mathematics education (Friel et al., 2001). When data are visualized in graphs, their meaning is not directly accessible, but must be inferred from the graph (Freedman and Shah, 2002). This involves relating the graphical information to a situational context (Leinhardt et al., 1990; van den Heuvel-Panhuizen, 2005). Previous research has revealed numerous types of errors and difficulties students encounter when interpreting graphs (e.g., Clement, 2001; Gagatsis and Shiakalli, 2004; Elia et al., 2007). Since both, mathematical aspects and the situational context, are relevant for the interpretation of graphs, this dialectic nature of functions may play a special role for the interpretation of eye movements.

To investigate students' work with functions, especially graphs, ET appears to be a promising research tool. Recent studies in the domain of functions have used ET to investigate the role of graphic properties and verbal information for graph interpretation (Kim et al., 2014) and have observed how students make transitions between mathematical representations (Andrá et al., 2015). These results indicate that ET has potential for investigating students' work with functions. This would link to the results from various other mathematical domains where the analysis of eye movements appears to be promising (Strohmaier et al., 2020; Schindler, 2021).

Since the interpretation of gaze data is not trivial and the EMH, on which many studies rely, has limitations, researchers need to know how ET can be applied in a certain domain. Thus, before one can pursue the long-term goal to validly apply ET in empirical studies focusing on the domain of functions, methodological studies are essential. Yet, to the best of our knowledge, researchers have not yet investigated how eye movements can be interpreted and to what extent the EMH can be applied in the domain of graph interpretation. Therefore, the aim of this paper is to investigate how eye movements can be interpreted in the domain of functions with respect to the EMH, in particular in students' interpretations of contextual graphs. Its aim is methodological in that it investigates the opportunities and challenges of ET. In particular, we ask the following research questions:

(1) Do students' eye movements correspond to their cognitive processes in the interpretation of graphs, and how?
(2) In how far are students' eye-movement patterns ambiguous or unambiguous?

Both research questions focus additionally on the role of the situational context, as it is our overarching goal to also get a better understanding of its impact for the relationship between eye movements and cognitive processes.

Our study builds on the work by Schindler and Lilienthal (2019) as they have investigated in the field of mathematics education research to what extent the EMH applies in the subdomain of geometry and how eye movements can be interpreted. On the one hand, our study connects to their study, as both deal with graphical forms of representation (hexagon and graph of a function, respectively). On the other hand, our study advances it with respect to the dimension of application. In contrast to purely inner-mathematical geometrical problems, we use graphs of contextual functions that relate to a real-world context. We examine these application-related contexts and their effects on the interpretation of eye movements in the subdomain of graph interpretation.

Schindler and Lilienthal (2019) conducted a case study using ET and stimulated recall interviews (SRIs), which illustrated that the interpretation of eye movements in this domain turns out to be challenging. Eye movements in geometry often cannot be interpreted unambiguously since mapping gaze patterns with cognitive or affective processes is not bijective. The design of our study was similar. In an exploratory case study, two university students worked on contextual graphs in three different situational contexts, wearing ET glasses. Directly after, SRIs were conducted with the students, in which they watched a gaze-overlaid video of their work on the tasks and recalled their thoughts while interpreting graphs. The analyzed data consist of the transcripts of all utterances from the SRI alongside with the eye movements from the work on the tasks.

Eye tracking and the eye-mind hypothesis in mathematics education research and beyond

Research on eye movements has considerably increased in recent years (König et al., 2016), also in mathematics education research. ET is used in many fields in mathematics education (e.g., numbers and arithmetic, reasoning and proof, and the use of representations), applying numerous methods to gather and analyze data (Strohmaier et al., 2020). In mathematics education research, a distinction can be made between studies that take an embodied perspective on eye movements (e.g., Abrahamson et al., 2015; Abrahamson and Bakker, 2016) and those that take a psychological perspective (e.g., András et al., 2015; Bruckmaier et al., 2019; Wu and Liu, 2022). Embodiment theories do not consider the mind and the body as separate entities since cognition is grounded in sensorimotor activity and, thus, eye movements are an integral part of cognition (Abrahamson and Bakker, 2016). The aim of using ET from a psychological perspective, in contrast, is to draw conclusions about cognitive processes by capturing eye movements. Here, eye movements are understood as a window to cognitive processes (König et al., 2016). Cognitive processes are defined as "any of the mental functions assumed to be involved in the acquisition, storage, interpretation, manipulation, transformation, and use of knowledge. These processes encompass such activities as attention, perception, learning, and problem solving" (American Psychological Association [APA], 2022). However, ET "is not mind reading" (Hannula, 2022, p. 30), but can provide insights into information processing, e.g., to "understand how internal processes of the mind and external stimuli play together" (Kliegl et al., 2006, p. 12). The gained insights can enrich the understanding of how learners acquire knowledge (Schunk, 1991). Indeed, Strohmaier et al. (2020) summarize studies that focus on "aspects of visualization" as well as those referring to "cognitive processes that cannot be consciously reported" (p. 167) as two of the three areas in mathematics education research where ET is particularly beneficial. Still, the relationship between eye movements and cognition is not immediately clear. Therefore, Schindler and Lilienthal (2019) emphasize that a discussion is needed about how ET data can be interpreted, i.e., to what extent the EMH applies, irrespective of what perspective (embodied or psychological) on eye movements one takes, in different mathematical domains.

Just and Carpenter (1976b) hypothesize that "the eye fixates the referent of the symbol currently being processed" (p. 139; EMH). According to this hypothesis, the eyes fixate the object that is currently on the top of attention. The EMH was derived from cognitive research dealing with reading (Just and Carpenter, 1976a). In addition to reading research, many researchers use them in other fields as the basis for their data analysis and interpretation. However, meanwhile it is known that these assumptions must be considered with caution, as there are some limitations. Even studies from the original field, i.e., reading research, have shown that information from previous fixated words and upcoming, not yet fixated, words also influence fixation durations of the current word (Underwood and Everatt, 1992; Kliegl et al., 2006). This "weakens the assumptions, because what is being fixated is not necessarily what is being processed" (Underwood and Everatt, 1992, p. 112). Moreover, there are situations in which words are processed by the participant although they are skipped, i.e., not fixated (Underwood and Everatt, 1992). Kliegl et al. (2006) summarize that the "complexity of the reading process quickly revealed serious limits of the eye-mind assumption" (p. 13). In a related area, the eye-mind hypothesis was challenged by observing eye movements during retrieval processes of read sentences. Again, the limitations of the assumptions became apparent: "Eye movements say nothing about the underlying retrieval process because the process controlling the switch in gazes is independent of the process controlling retrieval" (Anderson et al., 2004, p. 229). The fact that the EMH is also applied and investigated in completely different fields, namely in science, is shown in a study by Wu and Liu (2022) in scientific argumentation using multiple representations. They "examined the degree of consistency between eye-fixation data and verbalization to ascertain how and when the EMH applies in this subdomain of scientific argumentation" (Wu and Liu, 2022, p. 551) in order to contribute to reduce the ambiguity and clarify the validity of ET data in this subdomain. They conclude that "verbalizations and eye fixations did not necessarily reflect the same or similar cognitive processes" (Wu and Liu, 2022, pp. 562–563). They call for researchers to examine what factors have an influence on the relationship between eye-fixations and mental processes, i.e., on the EMH such as task/domain properties or prior knowledge. In mathematics education, Schindler and Lilienthal (2019) have illustrated in the subdomain of geometry that the EMH only partially holds true, which suggests that also in other mathematical subdomains the interpretation of ET data is not trivial either. One more reason for this is that eye movements do not only indicate cognitive processes, but also affective processes, i.e., processes characterized by emotional arousal, such as excitement about a discovery or panicking because of noticing a mistake (Schindler and Lilienthal, 2019). Hunt et al. (2014) found that mathematics anxiety affects (arithmetic) performance, as evidenced in significant positive correlations between math anxiety and gaze data, such as fixations, dwell-time, and saccades. Stress can also affect gaze behavior, as suggested by a study by Becker et al. (2022). Here, stress of mathematics teachers was indicated in the diagnosis of difficulty-generating task features and found to affect higher processing and related gaze behavior.

Therefore, the relationship between eye movements and cognitive (and affective) processes, i.e., how eye movements can be interpreted, needs to be investigated first. In the field of functions, this investigation is still missing. Thus, we take a first step in this direction and analyse eye movements and their interpretations for the subdomain of graph interpretation.

Eye tracking in the field of functions

Little ET research has been pursued in the domain of functions so far, so its potential for this domain is still unknown. Strohmaier et al. (2020) do not include the domain of functions as an own category in their survey; they classify some studies under the term "use of representations." Anderson et al. (2004), Andrá et al. (2015) compared experts and novices when focusing on different representations of functions. They showed that there are quantitative and qualitative differences between experts and novices that indicate that experts proceed more systematically than novices in terms of the order of looking at and considering the representations that may correspond to each other. Kim et al. (2014) investigated graph interpretation with line graphs and vertical/horizontal bar graphs of students with dyslexia. They measured reaction times and showed that the gap in reaction times between college students with and without dyslexia increases with the increasing difficulty of the graph and the question. Shvarts et al. (2014) investigated the localization of a target point in a Cartesian coordinate system and showed that experts have the ability to use additional essential information and to distinguish essential parts of visual representations, whereas novices often focused on irrelevant parts. In addition, there are studies referring to less typical representations of functions. Boels et al. (2019) studied strategies interpreting histograms and case-value plots. The most common strategies they found for students' interpretations of these graphs are a case-value plot interpretation strategy and a computational strategy. Reading values from linear versus radial graphs is the focus of Goldberg and Helfman's (2011) investigation that outlines three processing stages in reading values on graphs [(1) find dimension, (2) find associated datapoint, (3) get datapoint value].

In contrast to all these empirical studies, the focus of our ET study follows a novice approach in the domain of functions as it has a methodological focus. It investigates in what ways eye movements correspond to cognitive processes in students' interpretations of contextual graphs. In this subdomain, it is not yet clear how the situational context influences the relationship between eye movements and cognitive processing. We assume that diverse and complex cognitive processes accompany the occurrence of situational context as additional dimension. With our methodologically focused study, we hope to contribute to a better interpretation of the results in the domain of functions. As context is also relevant in other mathematical and scientific domains, our results might also be important for other mathematical domains in which context plays a role.

Graph interpretation in mathematics education

The concept of function is central in mathematics, regardless of the level at which mathematics is studied (e.g., Sajka, 2003; Doorman et al., 2012). While there is disagreement among mathematicians around which aspects of a function are crucial (Thompson and Carlson, 2017), what is understood as a function is less controversial. Dirichlet–Bourbaki authored a widely acknowledged definition, which Vinner and Dreyfus (1989) summarize as:

A correspondence between two non-empty sets that assigns to every element in the first set (the domain) exactly one element in the second set (the codomain). To avoid the term correspondence, one may talk about a set of ordered pairs that satisfies a certain condition (p. 357).

With regard to functions, three typical kinds of external representations can be distinguished: tabular, graphical, and algebraic (Sierpinska, 1992). Our work focuses on graphical representations, specifically on graphs, whose data originate from measured values of real-world situations. We call these graphs contextual graphs. These graphs are very common in daily life, but mostly underrepresented in school lessons. In school, the focus is rather on function types such as linear or exponential graphs. Nevertheless, contextual graphs are of central importance for the development of an ample mathematical literacy, especially with respect to the translation between real-world data and mathematical representations. In particular, recognizing and being able to interpret different representations of data belongs to statistical literacy and is central for the development of critical thinking (Garfield et al., 2010). Thus, learning how to deal with contextual graphs contributes to enable students to understand (media) reports using graphical representations and distinguish between credible and incredible information, interpret and critically evaluate them in order to use them as a basis for decision-making (Sharma, 2017).

Real-world contexts play a major role in interpreting contextual graphs. Graphs visualize data, which correspond to a functional context, such as the development of stock market prices, temperatures, or training processes. When data are visualized in graphs, their meaning is not directly accessible, but must be inferred from the graph (Freedman and Shah, 2002). Individuals have to make sense of the given information by processing it cognitively. To derive meaning from the information given in contextual graphs, it is necessary that the data must not only be extracted and understood, but also must be related to the situational context (Leinhardt et al., 1990; see e.g., van den Heuvel-Panhuizen, 2005 for contexts). Sierpinska (1992) points out that recontextualization (the reconstruction of the context from the given data, see van Oers, 1998, 2001) and the ability to relate given graphical information to a situational context together are the main difficulties that students face when interpreting graphs. This ability also depends on how much the setting of the graph is contextualized or abstract (Leinhardt et al., 1990).

When students interpret graphs, different levels of questions can be involved:

An elementary level focused on extracting data from a graph (i.e., locating, translating); an intermediate level characterized by interpolating and finding relationships in the data as shown on a graph (i.e., integrating, interpreting), and an advanced/overall level that requires extrapolating from the data and analyzing the relationships implicit in a graph (i.e., generating, predicting). At the third level, questions provoke students' understanding of the deep structure of the data presented (Friel et al., 2001, p. 130).

These levels of graph interpretation involve different cognitive processes, related to the perception and interpretation of graphs. It can be assumed that the analysis of students' eye movements can provide valuable insights into students' graph interpretation processes. This would suggest that ET is a valuable method to investigate how students interpret this kind of visual representations.

Materials and methods

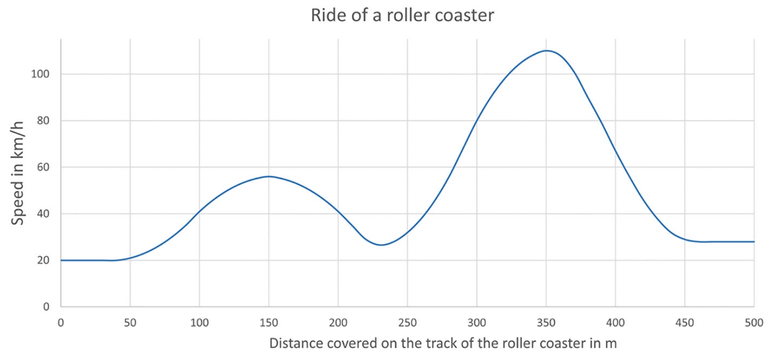
Participants, task design, and setting

We studied eye movements to investigate cognitive processes while working with functions represented as graphs. We conducted a case study with two university students. It can be assumed that they are more experienced in interpreting graphs than school students. In addition, they are more adept at reporting on their cognitive processes in SRIs, which was crucial for this exploratory study. We further chose two students with different backgrounds and affinity with respect to mathematics, in order to obtain a wider range of gaze patterns and approaches when working on the tasks. We chose two 21- and 28-years-old German university students. Gerrit (21) studied Engineering and Management with a focus on Production Engineering. He had a high affinity for and was interested in mathematics. Mathematics was a relevant domain in his professional field, so he was regularly occupied with mathematics at the time of study. In contrast, Elias (28) was an education student with a focus on German and history and did not have a specific interest in mathematics. At the time of study, he, therefore, was not used to work on mathematical problems. Both participants were communicative and volunteered to be participants in our study.

We presented graphs in three different situational contexts (**Figures 1–3**). They are inspired by the material published by the Shell Centre (1985): Two units deal with the change of velocity in a car race and a roller coaster ride, respectively and the third shows the change of the filling height of a vessel that is constantly being filled with water. Each unit consists of five tasks in which the participants were asked to interpret data from the graphs at different question levels, as specified by Friel et al. (2001; see **Figures 2, 3**). Each unit starts with an information slide about the situational context and the graph. This way, the participants have a chance to familiarize themselves with them. The first question asks them to describe the change of the velocity/filling height (intermediate level). The following two tasks ask them to extract information from the graph in the form of a single point (elementary level). Either a point on the abscissa is given to which the corresponding point on the ordinate must be found by reading information from the graph (task two), or *vice versa* (task three). In task four, the participants have to describe the change of the velocity/filling level in a specified interval (intermediate level) and interpret their result with regard to the situational context (overall level). Task five focuses on the interpretation of the whole graph, as the participants need to pick and justify the one out of four or five realistic images which they think represents the situation best (overall level).

The students worked on the tasks individually. The tasks were presented on a 24″ screen (60 Hz, viewing distance: ∼60 cm), each task on a single slide. There was no time restriction for working on the tasks. The students gave their

A roller coaster car is pulled up the first incline of the track and is released at a speed of 20 km/h. Then it drives without additional propulsion. The diagram shows the speed of the car on a section of the roller coaster from the first incline.



A vessel is filled with a consistent jet of water.

The diagram shows the water level in the vessel during filling.

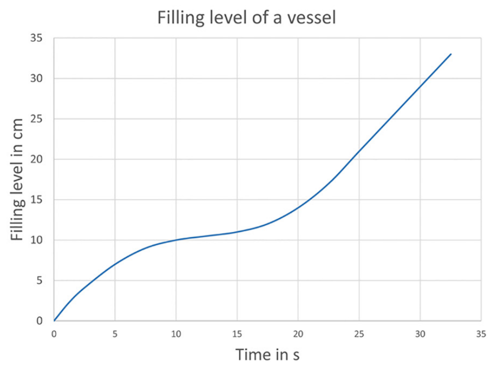


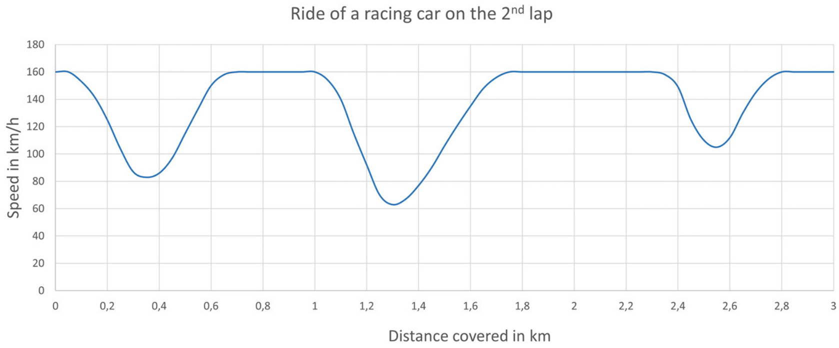
FIGURE 1
Situational contexts: change of velocity in a roller coaster ride, change of filling height in a vessel; translated to English.

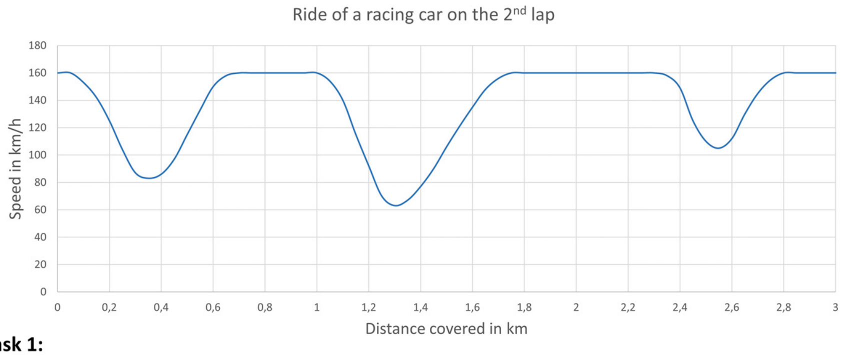
answers orally, while the sound was recorded by the built-in microphone of the ET glasses (see section "(Un)Ambiguity of eye movements patterns"). The students were able to get to the next task by autonomously clicking the computer mouse. The first author of this paper was the instructor of this work and was present during the students' individual work on the tasks. The instructor did not intervene while the students were working on the tasks unless the students directly addressed her and asked something. The instructor responded to the participants' questions in terms of task formulation, but did not provide any help with regard to the mathematical content.

Eye tracker and eye tracking data

To record eye movements, we decided to use a head-mounted system, as the integrated scene camera also records gestures and student utterances in a time-synchronized manner. Moreover, this system has the advantage that all data are synchronized and this does not have to be done subsequently, what is also important for the timely conduction of the SRIs. We used Tobii Pro Glasses 2 with 50 Hz. The binocular eye tracker allows tracking gazes through ET sensors and infrared illuminators. At the beginning of data collection, first, a single-point calibration was performed with the eye tracker to enable the transfer function that maps the gaze point onto the scene image. Then, an additional nine points calibration verification was performed, so that we could later check the measurement's accuracy. We repeated this verification procedure toward the end of the students' work to find possible deviations from the beginning. Gaze estimation under ideal conditions with Tobii Pro Glasses 2 is 0.62° (Tobii Pro AB, 2017). In our study, the average accuracy from the calibration at the beginning and the end of data collection was 1.1°, A race car drives its second lap on an even racetrack.

The diagram shows the speed while driving on the racetrack.

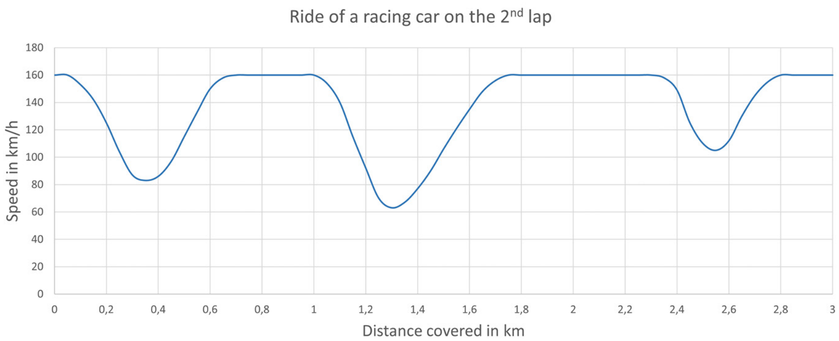




Task 1:

Describe how the velocity of the car changes while driving.

How can you tell from the graph?

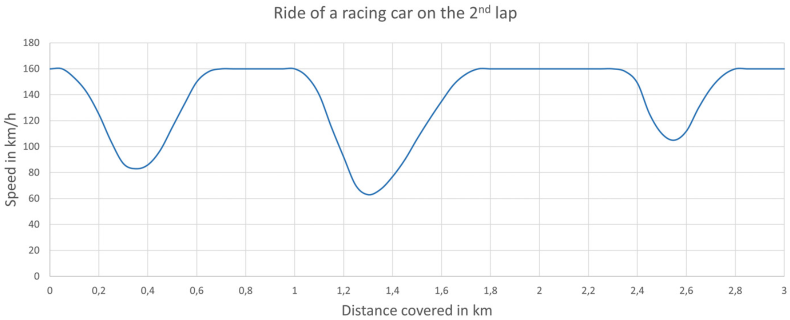


Task 2:

How fast does the car drives when it has covered exactly 0,6 km of the second lap?

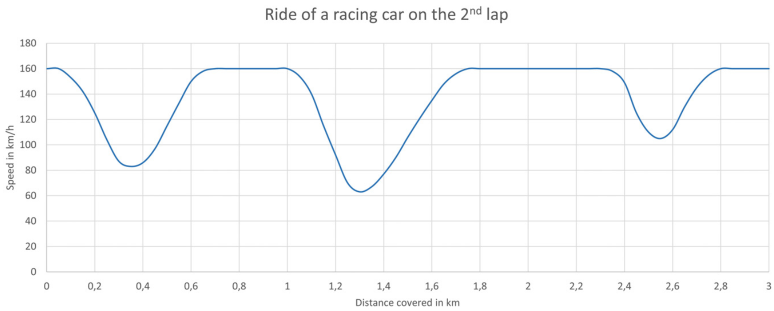
FIGURE 2
Introduction and tasks 1, 2 given to the students on the situational context: change of velocity in a car race; translated to English.

which corresponded to 1.15 cm on the screen we used. This inaccuracy was taken into account in the task design by placing all relevant task elements appropriately far enough apart. In addition, we also considered the inaccuracy in the course of data interpretation by not making any statements about situations in which it could not be clearly determined



Task 3:

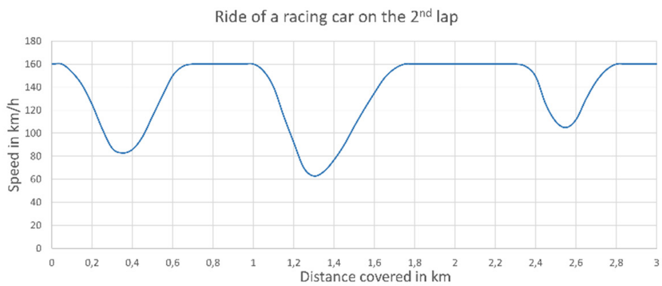
After how many kilometers does the car have a speed of exactly 80 km/h?

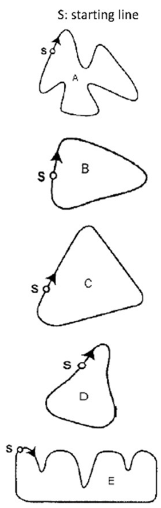


Task 4:

How does the speed of the roller coaster vary on the stretch between 2.3 and 2.8 km?

How is that related to the course of the racetrack?





Task 5:

Here you can see pictures of five racetracks. On which of these racetracks

did the car drive so that the shown speed graph occured?

Give reasons for your answer.

FIGURE 3
Tasks 3–5 given to the students on the situational context: change of velocity in a car race; translated to English.

what the person was looking at, despite the spacious task design.

According to Holmqvist and Andersson (2017), in only 2° of the humans' visual field, high-acuity vision is possible. This small area is called fovea. The ET method makes use of the small size of the fovea, since the eye must be directed to the area in which information is aimed to be extracted. Nevertheless, the surrounding part of the visual field, the large area of peripheral vision, is used for orientation. In it, the information processed is blurred and in black and white, but serves as an indication of the next target of the gaze, or to perceive movement in the periphery (Holmqvist and Andersson, 2017). Most eye movement studies predominantly analyse fixations (moments when the eye remains relatively still; approx. 200 ms up to some seconds) and saccades (quick eye movements between fixations; approx. 30–80 ms) (Holmqvist and Andersson, 2017). Hannula (2022) emphasizes that "different methods have been developed to analyse eye movement behavior. As in all areas of research, a phase of qualitative research has been necessary to get a basic understanding of the eye movements in a specific task" (p. 20). Therefore, our study focuses on a methodological research question, building on a qualitative approach. We investigate how eye movements can be interpreted in the domain of functions with respect to the EMH, particularly in students' interpretation of contextual graphs. Hence, we decided not to limit ourselves to certain measures (e.g., fixation durations, dwell time, or areas of interest), but to analyze raw data (Holmqvist and Andersson, 2017). Our purpose is not only to know where the student's attention is when interpreting contextual graphs, but also to interpret the eye movements themselves and relate them to cognitive processes. We, therefore, use eye movements as displayed in gaze-overlaid videos. The gaze-overlaid videos were produced by the software. See here Thomaneck et al. (2022) for an example. **Figure 4** shows an example of a successive gaze sequence as displayed in a gaze-overlaid video and **Figure 5** a merged visualization of this gaze sequence as a gaze plot, in which the order of fixations and the respective duration are displayed through the size of the corresponding circle.

Stimulated recall interview based on gaze-overlaid video

Lapses in memory often are a problem in traditional interviews. Dempsey (2010) expounds that "motivations and rationales that informants describe retrospectively may not conform to those they actually held in the moment of experience" (p. 349). SRI is a special form of interview in that participants are invited to recall their thinking during an event that is prompted by a stimulus. SRI is well suited to investigate cognitive processes (Lyle, 2003), since it "gives participants a chance to view themselves in action as a means to help them recall their thoughts of events as they occurred" (Nguyen et al., 2013, p. 2).

One kind of stimulus for this form of interview are gaze-overlaid-videos. These are videos that show the original scene, overlaid with gazes from the ET. Stickler and Shi (2017) point out that gaze-overlaid videos provide a strong stimulus for SRI, because this stimulus makes eye movements visible, which, for the participants, are usually not conscious. In our study, the gaze videos show the participants' processes of solving contextual graph interpretation tasks as a stimulus for recalling their cognitive processes while solving the tasks.

It is important to keep the time span short between ET and SRI and to ensure that the questions asked in the interview situation do not alter the cognitive processes that have taken place at the time of the event (Dempsey, 2010; Schindler and Lilienthal, 2019). In our study, the SRI was carried out by the first author about 30 min after the participants had completed the ET tasks. The SRIs lasted between 8 and 15 min per situational context. We used the video captured by the ET glasses, overlaid with the recorded gaze data and supplemented with the voice recording by the Tobii Controller Software (Tobii Pro AB, 2014) as stimulus. Prior to the SRI, the participants were told that they would see their eye movements in the form of a red circle. Then they were familiarized with the aim of the interview: that they should explain these gazes and explain their original thoughts to make their approaches understandable for the interviewer. The interviewees also wore the ET glasses during the SRI, so that verbal utterances and potential gestures could be recorded. The questions invited them to express their thoughts, or explain the rationales after having watched the stimulus. For example, we asked, *What did you do there?,* or *Why did you look so closely at this section of the graph at this moment?* Either the participants paused the video autonomously to explain their eye movements, or the interviewer herself stopped the recording and invited them to clarify the situation by asking a related question. The kind of questions was planned in advance to ensure that similar questions and wordings were chosen in both interviews. While the students were working on the tasks, the interviewer had the opportunity to observe the eye movements on a second screen and, thus, had time to consider at what moments she would pause the video if the students did not stop it themselves. Additional follow-up questions also arose spontaneously during the interview situations. Whether or not the interviewer posed follow-up questions, also depended on how detailed the students explained their eye movements.

In the SRIs, we found that both participants, Elias and Gerrit, were able to answer most of the questions and to comprehensibly explain their eye movements. Moreover, we strongly assume that their utterances actually correspond to their original cognitive processes, for the following reasons: It is known from psychological literature that the more deeply a stimulus has been analyzed, the better it can be recalled. Craik and Lockhart (1972) distinguish for instance between

10.3389/feduc.2022.1003740

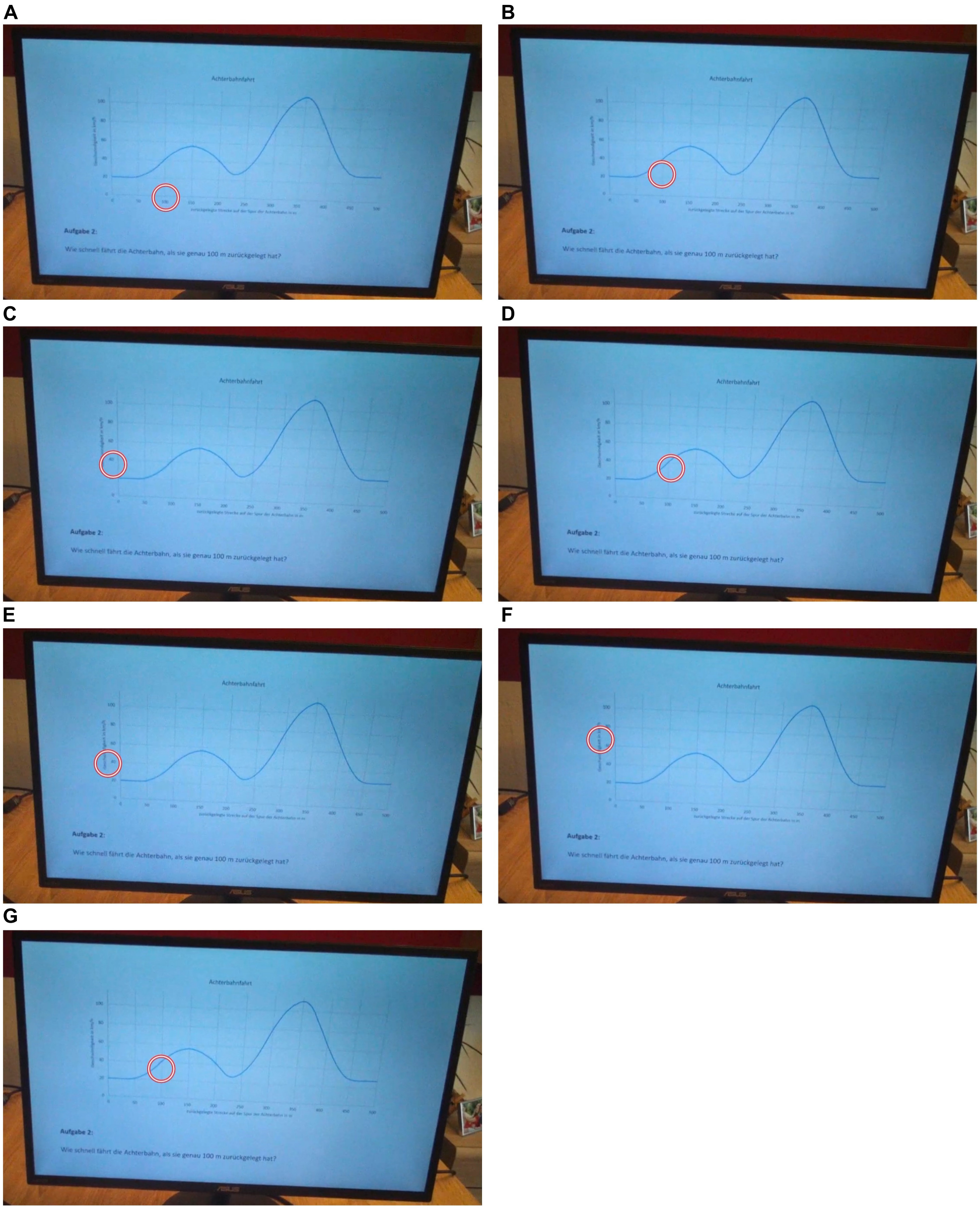


FIGURE 4
Gaze sequence **(A–G)** as displayed in a gaze-overlaid video.

preliminary stages of processing that are concerned with the analysis of physical or sensory features and later stages of processing that combine new input with previous knowledge and are concerned with pattern recognition and the extraction of meaning. In our study, a deep analysis of the stimulus by the participants can be assumed, since the tasks dealt with can be characterized by pattern recognition and the extraction of meaning and, thus, belong to later stages of processing. In addition, the students are reminded of their processing of the tasks by showing them their gaze-overlaid videos, which provide reflection-aiding. Nevertheless, in the few situations when they could not recall their original thoughts, they openly admitted it. This indicated that the gaze-overlaid video was a strong stimulus for the SRI in our study and the resulting utterances were a good data basis for our analyses. However, it should be noted here that the data is based primarily on what students say. Even though there are strong indications and arguments that these are credible and actually reflect their original thoughts, it cannot be entirely ruled out that these are complete and always true. Nevertheless, we consider the combination of ET and SRIs to be extremely helpful, as it allows us to get close to the thoughts and learning processes of the participants.

 10.3389/feduc.2022.1003740

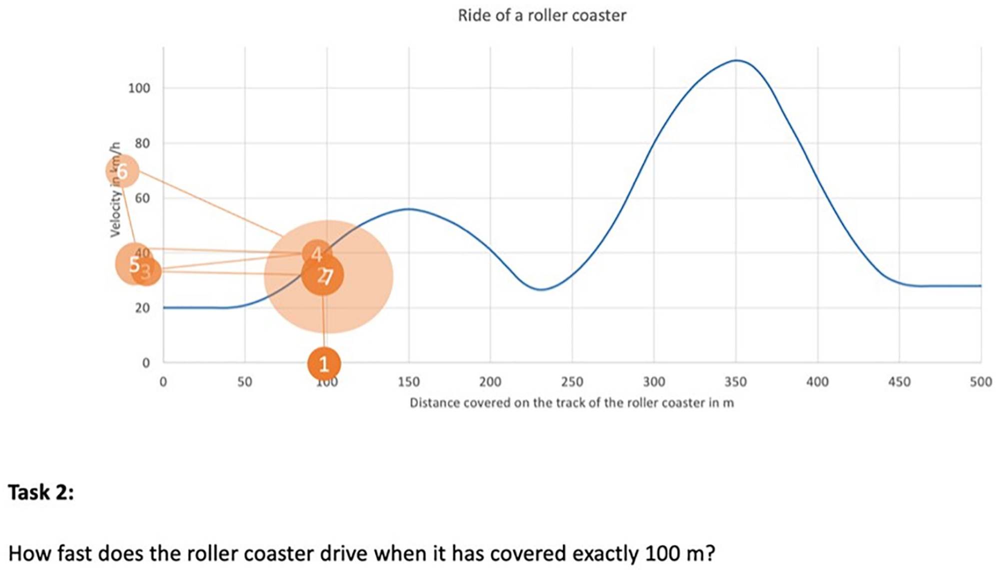


FIGURE 5
Gaze plot for the gaze sequence presented in **Figure 4**.

Data analysis

To analyze the data, we first transcribed the utterances from the ET videos, including the given answers to the tasks. The transcripts were organized in the first column of a table (see **Table 1**). The second column contained the transcription of gazes and utterances from the SRI video. When the video was paused in the SRI, the utterances and potential gestures of both the interviewer and participant were inserted in the corresponding line in the table. Thus, by presenting the eye movements and simultaneously spoken words on the same line (**Table 1**), we display which events happened concurrently. Afterward, we analyzed the SRI transcripts following Schindler and Lilienthal's (2019) adaption of Mayring's (2014) four steps of qualitative content analysis for ET-data (see data analysis steps in **Table 1** and the corresponding gaze sequence as displayed in a gaze-overlaid video in **Figure 4** and the corresponding gaze plot in **Figure 5**). We chose to use inductive category development, due to the explorative and descriptive nature of the research aim in this study.

The first step in Mayring's approach is the *transcription*, which, in our case, embraces both, the transcriptions of the utterances from the task processing, as well as the transcription of gazes and utterances from the SRI. In a first analytical step (Mayring: *paraphrase*), we paraphrased the elements from the transcripts with attention to relevance for our research interest, which include the eye movements and the interpretations given by the participants in the SRI. In the next transposing step (Mayring: *generalization*), the paraphrases were uniformed stylistically. Only after this, in the third step, did the actual development of categories take place (Mayring: *reduction*). Throughout the whole analytical process, we distinguished between gaze categories (shown in the upper part of the cells) and interpretation categories that describe the cognitive processes associated with the respective gazes (lower part of the cells following the colon). We carried out these steps with one third of the data to develop the category system. Then, the remaining data was organized using the preliminarily developed categories. Subsequently, we revised the category system by partially re-naming categories to unify the nomenclature. Furthermore, we arranged the categories in thematic main categories (e.g., gazes on the text, gazes on the graph, or recurring gaze sequences for the gaze categories).

When the interpretation was unclear and the assignment of the categories appeared to be uncertain, for example, because the interpretations given in the SRI were too imprecise or too general, we tried to verify our interpretation with the help of the utterances from the ET video in the left column. Finally, all interpretations that belonged to a certain gaze category were grouped to show all cognitive interpretations that matched a certain gaze pattern in our data. Examples of gaze patterns are gaze jumps between a point of the graph and the corresponding point of the axis or several gazes in succession at points of a graph section.

Results

In the following, we present the results with respect to the two research questions: (1) Do students' eye movements correspond to their cognitive processes in the interpretation of

TABLE 1 Example of data analysis steps (translated by the authors).

Data		Data analysis steps		
Transcript ET (Utterances)	**Transcript SRI (Gazes and utterances)**	**1st step: Paraphrase (Gazes and utterances)**	**2nd step: Transpose (Gazes and processes)**	**3rd step: Category (Gazes and processes)**
E: Slightly more than 40 km/h if the. . .	The gaze rests for 5 s at the point [100; 41] of the graph. I: Why are you looking up there for so long? E: Because that doesn't exactly hit the 40. That's why I wanted to see whether it was, whether it was just a mistake because I didn't see it correctly.	E. looks for 5 s at the point of the graph: E. says that the 40 is not hit exactly and he considers whether there is a mistake or he missed something.	Long fixation of a point of the graph: Reading off a point that is not at the level of an auxiliary line, which leads to difficulties.	Long fixation of a point of the graph. Reading off a point. Difficulties in processing the task.

TABLE 2 Correspondences of eye movements and cognitive processes.

	Eye movement pattern (Identified in the gaze-overlaid video)	**Cognitive/Affective processes (Described by the participants in the SRI)**
Correspondences	Gaze follows a line of text.	Reading of given information and data. Verifying given information and data. Matching the intended verbal answer to the task in terms of formulation.
	Fixations on turning points of the graph.	Focusing on prominent sections of the graph.
	Gaze jumps between corresponding parts of the graph and the realistic image.	Searching for similarities between graph and image. Detection of correspondences between graph and image. Detection of inconsistencies in graph and image. Reassurance of (partial) results in their own work on the task.
	Fixations on non-meaningful points.	Focusing on prominent sections of the graph or image (with peripheral vision). Quickly taking in the text, graph or image (with peripheral vision).
Differences	Quick saccadic eye movements.	Affective unsettledness. Uncertainty concerning mathematical aspects. Noticing a mistake.
	Fixations on non-meaningful points in the middle of the diagram.	Preparing/waiting for the next task.
Indirect correspondences	Gaze following the course of the graph.	Grasping the situational context. Grasping the course or properties of the graph. Imagining the realistic object or situation.

graphs, and how? (2) In how far are students' eye-movement patterns ambiguous or unambiguous? Table 2 (correspondences of eye movements and cognitive processes) and Table 3 (unambiguous eye movement patterns and associated cognitive processes) provide a summary of the main results presented in the following sections.

Correspondence of eye movements and cognitive processes

In ET research, a close relationship between eye movements and cognitive processes is often assumed. The underlying basic assumption is that what the eyes fixate on is cognitively processed at that very moment. We will present our results with respect to research question 1 below.

Correspondences of eye movements and cognitive processes

Our analyses suggest that students' eye movements were related to their cognitive processes in most instances. This applies to all elements of the stimuli (text, diagram, and graph). For instance, gazes that followed a line of text were explained in such a way that the text was read and understood. According to the participants, a fixation on the axis label served to grasp the meaning of an axis, whereas gazes on the marked points on the axes were used to orientate and find certain points (e.g., for reading a value or naming an interval). A gaze following the

TABLE 3 Unambiguous eye movement patterns and associated cognitive processes.

Eye movement (Identified in the gaze-overlaid video)	Cognitive processes (Described by the participants in the SRI)
Fixations on the labeling of an axis.	Taking in and understanding the meaning of the axes.
Fixation on the title of the diagram.	Taking in and understanding the meaning of the diagram.
Fixation on non-meaningful points in the center of the diagram.	Preparing/waiting for the next task.
Quick saccadic eye movements on different points of the stimulus or on non-meaningful points of the task sheet.	Emotional arousal due to feelings of insecurity, affective unsettledness, noticing a mistake, or difficulties in completing the task.

course of a graph section can indicate that the participant grasps the meaning and memorizes the course and certain properties of the graph. These are just a few of many examples, where the EMH held true, that is, where the cognitive processes were closely related to the students' eye movements.

Differences between eye movements and cognitive processes

Although the EMH held true in many instances, we also observed several situations in which the eye movements were not aligned with the students' cognitive processes. The gaze pattern where it was most obvious that the EMH did not hold true, was that of quick gaze jumps on different points of the (digital) task sheet – even semantically non-meaningful ones, i.e., areas from which no relevant information can be extracted to solve the task. The participants explained this eye movement pattern in different ways. What all of these instances had in common was that the places, where the students looked at, were not semantically related to the students' cognitive processing. Instead, emotional arousal dominated when this eye movement pattern occurred. For example, Elias described that he was in a state of affective unsettledness. He mistrusted the task and could not cope with the realistic image given as part of the fifth task. When the interviewer asked him about his quick eye movements in this situation in the SRI, he explained, "I felt tricked because I thought this is really one-to-one (chuckling) the same. [...] Because I'm just completely insecure at that moment and thinking I missed something." (All quotations from the data are translated from German by the authors.) We made similar observations in situations when the students noticed a mistake in their own answer, or felt insecure about mathematical aspects of the tasks. In all these instances, the eye movements appeared to be related to affective arousal. Thus, in addition to cognitive processes, affective processes are also reflected in the eye movements. Although the affective processes are related to the cognitive processes taking place, the EMH does not hold here, because the eye movements reflect the affective processes prevailing at that moment.

Another interesting eye movement pattern, in which the EMH did not hold true, occurred several times with both participants at the end of a subtask. After finishing and responding to a task, their eyes fixated on non-meaningful points in the middle of the diagram on the digital task sheet. However, no relevant information to solve the task can be extracted here, so these points were semantically not meaningful. Elias explained that he looked at this point, because he "immediately knew that directly after this, the next will appear right away, the next page." This eye movement, thus, had no reference point in terms of the functional context or the requirements of the task.

Additionally, there were situations in which participants seemed to focus on a single point on the task. The SRI, however, revealed that they did not actually semantically process the information displayed at the fixated point, but, rather, perceived the surrounding area with peripheral vision. The peripheral area is the region of vision outside the point of fixation, in which we cannot see sharply (see section "Eye tracker and eye tracking data"). In our data, it sometimes seemed as if the participants fixated on a non-meaningful point on the slide, but actually covered a larger area of the stimulus using their field of vision. For instance, referring to the ET video, Gerrit focused several times on a point slightly above a minimal turning point or a little below a maximum turning point. However, the respective turning point is included in the area of peripheral vision. By calculating the ET accuracy (see section "Eye tracker and eye tracking data"), we could ensure that there was no technical error in the measurement. When the interviewer asked him in the SRI what he was doing there, he replied: "Eh, there I looked at the peaks of the graph again." Thus, he confirmed that he was cognitively processing a larger area, the peaks of the graph, and that the point he was fixating on was not identical to the focus of his thoughts. A similar situation appeared, for example, when the participants fixated on a single point of the graph or the realistic image, but actually covered a larger section of it or perceived the object as a whole, for example, to determine whether it changed in contrast to the previous task. Thus, the students did not cognitively process exactly what they fixated on, but, with the help of peripheral vision, a larger area of their field of vision. Considering this, the EMH does hold true again, though it does in a broader sense. As peripheral vision can capture information in every ET study, independent from the mathematical content, this is a domain-general challenge for the interpretation of eye movements.

Indirect correspondences of eye movements and cognitive processes: Graph vs. context

We observed instances where it was ambiguous whether the EMH held true or not. The ambiguous cases referred to situations in which the participants fixated on a certain object of the stimulus, but the cognitive processes that they described in the SRI did not directly relate to this object, but to the situational context of the task. In the SRIs, it became clear that when the students looked at the graph, their thoughts were often related to the situation and the real objects. For instance, in the task presenting the car race, Gerrit fixated on several points of the graph, partially jumping between the axis and the graph. He explained, "Eh, I looked for when the speed decreases. And that should actually stand for that you brake, like driving into a curve. And then the speed, of course, increases again. This means that you are through the curve and can accelerate again." Here, Gerrit described cognitive processes aligning with the situational context. In another case, Elias described in the SRI for the same task, "I try [. . .] to complete this image by simply figuring out, how much distance does it drive? And when could the curve appear? How long is the curve? [. . .]. That's why the curve must have a certain shape." Elias, when gazing at the graph, was imagining the race course and trying to capture the situation as precisely as possible with the aid of the graph. We found similar instances with eye movements that followed the course of the graph, while the participants were thinking about the situational context. These cases can be described as an implicit support of the EMH conditioned by the additional dimension of the situational context: Although the gazes related exclusively to the graph, the cognitive processes also partly related to the situational context. Thus, this challenge is specific for the domain of functions, in particular for the interpretation of contextual graphs, that is, graphs of functions that are linked to data from real-world situations.

(Un)Ambiguity of eye movements patterns

As indicated in the previous section, certain eye movement patterns can relate to different cognitive processes (ambiguous eye movement patterns), whereas others appear to always relate to the same cognitive process (unambiguous eye movement patterns). We will present the detailed results with respect to research question 2 below.

Unambiguous eye movements patterns

We found four eye movement patterns that were related to a singular cognitive process, reported by the students in the SRIs (**Table 3**). The first two relate to elements of the graph (labeling of axes and title of the diagram), which the students interpreted (grasp and understand the elements' meaning). The third eye movement pattern was related to processing the general task in our study: When the students' gazes rested on non-meaningful points in the center of the task sheet, after the participants had finished a subtask, they were preparing/waiting for the next subtask to come on the next slide.

In addition, we found an unambiguous eye movement pattern, where the eye movements were related to affective processes (see section "Differences between eye movements and cognitive processes"). Quick saccadic eye movements on different points of the stimulus, or on non-meaningful points of the task sheet, where the gaze wandered "hectically" over the task sheet, were related to emotional arousal. For instance, the pattern appeared over a period of 14 s when Elias was in a state of affective unsettledness. He described, "I'm just completely insecure at that moment and think I missed something." Other examples for quick saccadic eye movements occurred in situations in which emotional arousal related to processing the task, such as if the participant felt insecure about the task concerning mathematical aspects, noticed a mistake in his earlier work, or had difficulties completing the task. All these instances have in common that the eye movement pattern indicates emotional arousal.

Ambiguous eye movement patterns

To illustrate, in which situations the interpretation of the gaze movements were ambiguous, we first give some examples before moving on to a more general observation regarding this aspect.

A frequent eye movement pattern that was associated with different cognitive processes was presented in **Table 2**: The gaze fixated a point on one axis, moved to the corresponding point on the graph, and then moved on to the particular point on the other axis. One interpretation of this gaze pattern given by the participants was reading off the point of the graph. More specifically, they describe how they searched for the given point on the ordinate, then moved to the corresponding point in the graph, and, finally, read the value on the abscissa. Gerrit, for example, describes his approach to this in the context of the process of filling a vessel: "I looked and searched for the point of 20 s on the x-axis and then read the corresponding y-value." The second interpretation given in the SRI was that the students wanted to reassure themselves regarding their (partial) results.

A second example of an ambiguous eye movement pattern was following the course of a graph section with the gaze. On the one hand, the students reported that they wanted to grasp or memorize the course of the graph. On the other hand, the students tried to grasp and understand the situational context of the graph (see section "Indirect correspondences of eye movements and cognitive processes: graph vs. context"). For instance, Gerrit described, "Because I've tried to imagine how the filling level changes when the graph increases or when it no longer increases that much." Thus, on the one hand, there are interpretations of eye movements on the graph referring to the graph itself and, on the other hand, interpretations referring to the situational context.

The third example is a gaze pattern that occurs exclusively in the last task of each situational context, in which the students have to decide which realistic image is appropriate for the graph.

The gazes jump between (corresponding) parts of the graph and the realistic image. This gaze pattern indicates a wide range of cognitive processes, such as comparing graph and image, building figurative imagination, recognizing correspondences between graph and image, recognizing discrepancies between graph and image, excluding an image.

Overall, it can be seen that the degree of ambiguity varies for the different gaze patterns. There are gaze patterns that are almost unambiguous, such as the example given here first. The participants exclusively interpret this gaze pattern in the two ways mentioned above [reading off a point vs. reassuring of (partial) results]. Next, there are gaze patterns where there are somewhat more possibilities for interpretation, such as following the graph with the gaze (second example) to grasp or memorize the graph vs. grasp and understand the situational context related to the graph. There are remarkably many relating cognitive processes (e.g., comparing graph and image, building figurative imagination, recognizing correspondences between graph and image, recognizing discrepancies between graph and image, excluding an image) to the gaze pattern of gaze jumps between (corresponding) parts of the graph and the realistic image, what was our third example in this section.

Discussion

The aim of this article was to investigate how eye movements can be interpreted in the domain of functions with respect to the EMH, particularly in students' interpretations of contextual graphs. We conducted an exploratory case study with two students, in which we investigated (1) if students' eye movements correspond to their cognitive processes in the interpretation of graphs, and how and (2) in how far students' eye-movement patterns are ambiguous or unambiguous. We used ET together with SRIs and the cases of two university students with different proficiency to investigate these methodological lines of inquiry.

We found that the students' eye movements often corresponded to their cognitive processes. This suggests that studying the eye movements of students interpreting contextual graphs can provide researchers with insights into the students' cognitive processes, which confirms the potential that ET has in this domain. However, in some instances, the eye movements tracked to elements on the digital task sheet that had little to do with the associated cognitive processes. Besides cognitive processes, thanks to our bottom-up coding procedure, we were also able to find affective processes that participants gave as explanations for their eye movements. For example, we found that quick gaze jumps indicate affective arousal and that the fixation of a non-meaningful point of the slide at the end of a subtask can be interpreted as students preparing or waiting for the subsequent task. Furthermore, we have encountered a domain-general phenomenon, namely that the appearance of a fixation of a single point can also indicate peripheral vision, by means of which a surrounding area is perceived. Particularly interesting is the case, where cognitive processes correspond indirectly with eye movements. The students look at the graph, but perceive or imagine the situational context of the task that is caused by the additional dimension of this domain. When investigating whether the eye movement patterns were ambiguous or unambiguous, we identified that some eye movement patterns were related to a singular cognitive process, while we found that others had different associated cognitive processes.

Taken together, these results indicate that it is valuable to analyse eye movements in students' interpretations of graphs, since they relate closely to students' cognitive processes, and ambiguities are limited. However, to know exactly what the eye movements indicate, one needs additional information from the SRI in some situations, since, even with using ET, it is not possible to read the student's minds (Hannula, 2022). It seemed to be relatively apparent when the students were thinking about something other than the task, or were emotionally unsettled, so that their eye movements were no longer related to the semantics of the displayed stimuli on the task sheet. The fact that eye movements relate to both cognitive and affective processes confirms the findings from studies in other subdomains, in which both kinds of processes were found (Hunt et al., 2014; Schindler and Lilienthal, 2019; Becker et al., 2022).

Generally, our findings about the relationship of eye movements and cognitive processes are partially domain-general and partially domain-specific for functions and the interpretation of contextual graphs. One domain-general finding provided insight into students' readings of text. According to the participants, following a line of text with the gaze could be assigned to the cognitive process of reading and understanding the text. However, it is important to distinguish whether the text was read for the first time for information retrieval, or whether it was re-read in the further course of task processing to verify the information and data provided, or to match the intended verbal answer with the task in terms of formulation. Reading the text is only one example of the occurrence of the phenomenon that the same kind of eye movements can relate to different cognitive processes depending on the phase of task processing the student is in. There are similar findings with other eye movements. Based on our results, we hypothesize that there are at least three distinct phases in processing graph interpretation tasks: (1) initial orientation, in which an overview of the task and representations is obtained; (2) carrying out an approach to solve the task; (3) checking with respect to one's own results or/and in relation to the formulation of the task. This can be seen, for example, in the second task, in which a value has to be read off the y-axis. For example, the three phases have the following form when Elias works on the roller coaster context: First, in the initial orientation, the gaze follows the line of text twice. Here, he reads and understands the task. Then, his gaze jumps to several points on the two axes, to the heading of the diagram, as well as to the labels of both axes, to a point on the graph and again to the label of the x-axis. According to the information from the SRI, he gets an overview of the diagram, checks whether it is the same as in the previous task and makes sure which quantities are applied to the axes. Then the second phase begins, in which he carries out his solution process. His gaze fixates the point on the x-axis that is indicated in the text, jumps to the corresponding point on the graph and to the corresponding value on the y-axis. This is where he actually reads off the value. In the third phase, the checking of one's results with relation to the task, his gaze jumps several times between the relevant point on the graph and its value on the y-axis and then rests a little longer on the point of the graph. According to the SRI, this serves to assure himself of his own result. Then, in order to be able to give a suitable answer to the task that is also adapted to its formulation, his gaze jumps back again to the labels of the y-axis and the line of text. Elias' eye movements are an exemplary sequence of gazes in the respective phases. Our empirical data show that these gazes are often additionally enriched with supplementary gaze repetitions, e.g., reading off a certain point again and again. Even in this study with a very small sample, however, it became clear that these three phases do not occur with every person and every task. In routine tasks, for instance, some phases seem to be very short or even omitted. Yet, in more complex tasks, in which the approach is not clear from the outset, it is indicated that there may be a further phase, in which one considers how one could solve the task. This additional phase, however, could also be understood as an extended initial orientation phase. Moreover, there may be further additional phases that could not be observed in this study. However, the phases that could be observed in our study are very similar to the famous four-phase model of problem solving (Pólya, 1945) – (i) understanding the problem, (ii) devising a plan, (iii) carrying out the plan, and (iv) looking back – or more recent versions of this first version of this model. Still, they also differ in some nuances from the phases we observed, since our tasks (apart from one) were no typical problem-solving tasks. For instance, the second task, which served as illustrating example above, does not pose a problem for the students, but is a routine task. In this task, there was no need to search for a solution approach, so that only three phases could be observed here. However, the fact that the respective phase of task processing has an influence on the interpretation of eye movements should be taken into account in further studies. Therefore, further studies are needed to determine in which subdomains and for which task types such phases occur, and which eye movements and aligning cognitive processes emerge. In this regard, Schindler and Lilienthal (2020) studied a student's creative process when solving a multiple solution task. They identified similar phases: (1) looking for a start; (2) idea/intuition; (3) working further step-by-step (including verification, finding another approach, finding mistakes in previous approach, correcting the old approach); (4) finding a solution (including verification) or discarding the approach (Schindler and Lilienthal, 2020). Even if the task type used in their study is very different from those of this study and there are steps that do not appear in this study (e.g., intuition; finding another approach), some elements are comparable (e.g., an initial phase: orientation vs. looking for a start; verification processes). In particular, the findings of Schindler and Lilienthal (2020) for creative processes in multiple solution tasks illustrate that ET is beneficial for observing phases in task processing in detail. Therefore, it can be assumed that ET could also be an appropriate tool to identify phases in other fields or further specify existing phase models, e.g., of problem solving.

As another domain-general result, we found that quick saccadic eye movements, where the gaze was "hectically" wandering around on the task sheet, indicated affective unsettledness or other kinds of emotional arousal. For the domain of geometry, Schindler and Lilienthal (2019) found very similar results. In further studies in other (mathematical) domains, this observation should be further examined and verified, and when indicated, a domain-general theory of these kinds of eye movements and aligning affective processes in task processing could be developed. Wu and Liu (2022) call for the higher goal of scientific research "to formulate general rules that describe the applicability of the EMH under various subdomains" (p. 567). We consider our findings regarding the consideration of the phase of task processing and the appearance of affective processes as an initial step toward these general rules.

In addition to these domain-general findings in task processing, there are also numerous eye movements and aligning processes that relate directly to the interpretation of graphs. According to Friel et al. (2001), graph interpretation tasks can be assigned to elementary, intermediate, or overall level (see section "Introduction"). Elementary tasks require the extraction of information from the data. We implemented this in our task design as the reading of certain points from the graph. We found that one gaze pattern occurred repeatedly in these tasks and, thus, seems to be typical for this question level: The gaze fixated on a point on one axis, then moved to the corresponding point on the graph, and then moved on to the particular point on the other axis (see **Table 2**; this pattern was also found by Goldberg and Helfman, 2011). Similar to reading the text, these eye movements also occurred again in a later phase of processing the task, when the students wanted to validate the points or results. Here, again, ambiguity of the interpretation can be minimized by including the phase of task processing into the evaluation. However, one must keep in mind that the university students in our study knew how to read points of a graph; they were already proficient at elementary-level tasks. Whether eye movements can be (almost) unambiguously interpreted if the participants do not yet master this process remains an open question. We assume that different eye movement patterns might be possible.

The interpretation of eye movements, in our study, is less clear for intermediate and overall level tasks. We found that in tasks of these levels, there are other typical eye movements (e.g., intermediate level: following the line of the graphs; overall level: gaze jumps between parts of the graph and the realistic image). These gaze patterns have a higher, respectively a very high degree of ambiguity (see section "Ambiguous eye movement patterns"), since complex cognitive processes are included and there are interpretations referring to the graph, but also those referring to the situational context. This is because when the level of questioning increases, a reference to the situational context, in which the graph is embedded, must be established (Leinhardt et al., 1990; Sierpinska, 1992), which is much more demanding. The graphs in our tasks are contextualized (van Oers, 1998, 2001) since the data originate from a real-world situation and are represented in abstract contextual graphs. The students, when interpreting the graph in higher-level tasks, need to recontextualize the graph. Our findings indicate that the higher the task level is the more ambiguity is involved in the interpretation of eye movement patterns. Wu and Liu (2022) identified that the EMH holds as long as the information in the stimuli can be easily identified and does not necessarily hold when the information is less precise. Our findings specify this as they show that the degree of ambiguity of the interpretation of eye movements increases when the information of the representations are more difficult to extract, e.g., because of the task requirement to link the information from the graph to the situational context. In further studies, these processes and relationships should be investigated in more detail – also to examine to what extent this relationship between ambiguity and task level is a general rule that can be extended to other subdomains. This would be conceivable, for example, in the field of geometry. The study by Schindler and Lilienthal (2019) uses an inner-mathematical problem with a regular hexagon as a task. However, one could also set application-related tasks here, for example, on tiling a surface with hexagonal tiles, which would add an additional dimension – similar to the situational context in the graphs we used – which could cause ambiguity with respect to the interpretation of eye movements.

In addition to this domain-specific challenge in interpreting eye movements, we also found a domain-general challenge related to the specifics of ET methodology: The eye tracker suggests foveal vision since it is not able to display peripheral vision (Holmqvist and Andersson, 2017). Yet, in line with Kliegl et al. (2006), who observed parafoveal processing of previous and upcoming words in reading research, and Schindler and Lilienthal (2019) in mathematics education, we found that students use peripheral vision to perceive information when processing tasks by capturing larger areas than just a single focal point. Without additional SRI and only relying on ET, we would have made partially incorrect assumptions regarding the interpretation of eye movements in some instances. This further illustrates how carefully the interpretation of ET data must be handled.

Before we summarize this study and its contributions, we want to mention some limitations. The study was conducted with only two university students. However, since this is an exploratory study, the results are still meaningful and provide a good starting point for further studies that should, for example, clarify the transferability of the results to (secondary) students. Due to the methodological focus of the study, it was nevertheless possible to gain relevant insights into the interpretation of eye movements in students' interpretations of graphs. Furthermore, it is important to note that for investigating the correspondences between eye movements and according cognitive processes, we could not directly access students' cognitive processes, but used SRIs to gain insights into them. Based on the gaze-overlaid videos, the students recalled their original thoughts during their work on the tasks and reported their cognitive processes. This means that our interpretations were influenced by what the students reported. However, the two university students were well able to report and recall their thoughts in the SRIs directly after the original work on the tasks, which provided us with valuable and interesting insights.

In summary, we see high potential and benefits in using ET for the detailed study of students' processes in the interpretation of graphs. Our findings show that this requires a very careful approach to the interpretation of eye movements, i.e., to the application of the EMH. Our study contributes to increase the validity of using ET in the domain of graph interpretation, but also in other domains. We found that domain-general aspects, such as peripheral vision, influence the interpretation of eye movements. Therefore, when conducting ET studies, one must always keep in mind that this type of visual information acquisition cannot be captured by technology. Study directors must consider that it is not always (only) the area specified by the visualized viewpoint that is acquired, but possibly the surrounding area as well. In addition, different phases of task processing might play a role for the interpretation of eye movements with relation to cognitive processes, as they can cause ambiguity. Often, however, there are related cognitive processes to the eye movements that merely serve a different intention due to different phases of task processing (see section "Ambiguous eye movement patterns"), so that the ambiguity can be minimized by including the phase of task processing in the interpretation of eye movements. This can occur in many types of tasks and, thus, domains where multi-step processes are used to arrive at a solution. Our findings, moreover, suggest that the task level influences the degree of ambiguity. In the interpretation of contextual graphs, the additional dimension of situational context causes this partially. However, we cannot rule out the possibility that in other domains there may be other influencing factors that have a similar effect. Therefore, when interpreting ET data, it is necessary to check in each case, whether such an additional dimension, in the form of a situational context or of some other kind, exists. To conclude, with our methodological study, we hope to increase the validity of ET studies in the domain of functions and in particular, the interpretation of (contextual) graphs, but also to be able to give clues for other domains on how eye movements can be interpreted and what influencing factors they may have.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AT, MV, and MS contributed to conception and design of the study and wrote the manuscript. AT conducted the study, collected the data, and performed the data analysis. All authors read and approved the submitted version.

Funding

This publication fee was covered by the State and University Library Bremen.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abrahamson, D., and Bakker, A. (2016). Making sense of movement in embodied design for mathematics learning. *Cogn. Res. Princ. Implic.* 1:33. doi: 10.1186/s41235-016-0034-3

Abrahamson, D., Shayan, S., Bakker, A., and van der Schaaf, M. (2015). Eye-tracking piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Hum. Dev.* 58, 218–244. doi: 10.1159/000443153

American Psychological Association [APA] (2022). *Cognitive processes. APA dictionary of psychology*. Available online at: https://dictionary.apa.org/cognitive-process (accessed on Oct 25, 2022).

Anderson, J. R., Bothell, D., and Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychol. Sci.* 15, 225–231. doi: 10.1111/j.0956-7976.2004.00656.x

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., and Sabena, C. (2015). Reading mathematics representations: An eye-tracking study. *Int. J. Sci. Math. Educ.* 13, 237–259. doi: 10.1007/s10763-013-9484-y

Becker, S., Spinath, B., Ditzen, B., and Dörfler, T. (2022). Den stress im Blick – lokale Blickbewegungsmaße bei der Einschätzung schwierigkeitsgenerierender merkmale von mathematischen textaufgaben unter stress. *J. Math. Didakt.* doi: 10.1007/s13138-022-00209-7

Boels, L., Bakker, A., Van Dooren, W., and Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educ. Res. Rev.* 28, 1–23. doi: 10.1016/j.edurev.2019.100291

Bruckmaier, G., Binder, K., Krauss, S., and Kufner, H.-M. (2019). An eye-tracking study of statistical reasoning with tree diagrams and 2 × 2 tables. *Front. Psychol.* 10:632. doi: 10.3389/fpsyg.2019.00632

Clement, L. L. (2001). What do students really know about functions? *NCTM* 94, 745–748. doi: 10.5951/MT.94.9.0745

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *J. Verbal Learn Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Dempsey, N. P. (2010). Stimulated recall interviews in ethnography. *Qual. Sociol.* 33, 349–367. doi: 10.1007/s11133-010-9157-x

Doorman, M., Drijvers, P., Gravemeijer, K., Boon, P., and Reed, H. (2012). Tool use and the development of the function condept: From repeated calculations to functional thinking. *Int. J. Sci. Math. Educ.* 10, 1243–1267. doi: 10.1007/s10763-012-9329-0

Elia, I., Panaoura, A., Eracleous, A., and Gagatsis, A. (2007). Relations between secondary pupils' conceptions about functions and problem solving in different representations. *Int. J. Sci. Math. Educ.* 5, 533–556. doi: 10.1007/s10763-006-9054-7

Freedman, E. G., and Shah, P. (2002). "Toward a model of knowledge-based graph comprehension," in *diagrammatic representation and inference*, eds M. Hegarty, B. Meyer, and N. H. Narayanan (Berlin, HD: Springer), 18–30. doi: 10.1007/3-540-46037-3_3

Friel, S. N., Curcio, F. R., and Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *J. Res. Math. Educ.* 32, 124–158. doi: 10.2307/749671

Gagatsis, A., and Shiakalli, M. (2004). Ability to translate from one representation of the concept of function to another and mathematical problem solving. *Educ. Psychol.* 24, 645–657. doi: 10.1080/0144341042000262953

Garfield, J., del Mas, R., and Zieffler, A. (2010). "Assessing important learning outcomes in introductory tertiary statistics courses," In *Assessment methods in statistical education: An international Perspective*, eds P. Bidgood, N. Hunt, and F. Jolliffe (New York, NY: Wiley), 75–86. doi: 10.1002/9780470710470

Goldberg, J., and Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Inf. Vis.* 10, 182–195. doi: 10.1177/1473871611406623

Hannula, M. S. (2022). "Explorations on visual attention during collaborative problem solving," in *proceedings of the 45th conference of the international group for the psychology of mathematics education*, Vol. 1, (dialnet), 19–34.

Holmqvist, P. K., and Andersson, D. R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures*, 2. Edn. Lund Eye-Tracking Research Institute.

Hunt, T. E., Clark-Carter, D., and Sheffield, D. (2014). Exploring the relationship between mathematics anxiety and performance: An eye-tracking approach. *Appl. Cogn. Psychol.* 29, 226–231. doi: 10.1002/acp.3099

Just, M. A., and Carpenter, P. A. (1976b). Eye fixations and cognitive processes. *Psychol. Rev.* 87, 329–354. doi: 10.1016/0010-0285(76)90015-3

Just, M. A., and Carpenter, P. A. (1976a). The role of eye-fixation research in cognitive psychology. *Behav. Res. Methods* 8, 139–143. doi: 10.3758/BF03201761

Kim, S., Lombardino, L. J., Cowles, W., and Altmann, L. J. (2014). Investigating graph comprehension in students with dyslexia: An eye tracking study. *Res. Dev. Disabil.* 35, 1609–1622. doi: 10.1016/j.ridd.2014.03.043

Kliegl, R., Nuthmann, A., and Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *J. Exp. Psychol. Gen.* 135, 12–35. doi: 10.1037/0096-3445.135.1.12

König, P., Wilming, N., Kietzmann, T. C., Ossandón, J. P., Onat, S., Ehinger, B. V., et al. (2016). Eye movements as a window to cognitive processes. *J. Eye Mov. Res.* 9, 1–16.

Leinhardt, G., Zaslavsky, O., and Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Rev. Educ. Res.* 60, 1–64. doi: 10.3102/00346543060001001

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *Br. Educ. Res. J.* 29, 861–878. doi: 10.1080/0141192032000137349

Nguyen, N. T., McFadden, A., Tangen, D. D., and Beutel, D. D. (2013). "Video-stimulated recall intervies in qualitative research," in *joint australian association for research in education conference of the australian association for research in education (AARE)*, (Australian Association for Research in Education), 1–10.

Pólya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

Sajka, M. (2003). A secondary school student's understanding of the concept of function—A case study. *Educ. Stud. Math.* 15, 229–254. doi: 10.1023/A:1026033415747

Schindler, M. (2021). "Eye-Tracking in der mathematikdidaktischen Forschung: Chancen und Herausforderungen," in *vorträge auf der 55. tagung für didaktik der mathematik - jahrestagung der gesellschaft für didaktik der mathematik vom 01. märz 2021 bis 25. märz 2021*, (Oldenburg: Gesellschaft für Didaktik der Mathematik), doi: 10.17877/DE290R-22326

Schindler, M., and Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educ. Stud. Math.* 101, 123–139. doi: 10.1007/s10649-019-9878-z

Schindler, M., and Lilienthal, A. J. (2020). Students' creative process in mathematics: Insights from eye-tracking-stimulated-recall interviews on students' work on multiple solution tasks. *Int. J. Sci. Math. Educ.* 18, 1565–1586. doi: 10.1007/s10763-019-10033-0

Schunk, D. H. (1991). *Learning theories: An educational perspective*. New York, NY: Macmillan Publishing Co, Inc.

Sharma, S. (2017). Definitions and models of statistical literacy: A literature review. *Open Rev. Educ. Res.* 4, 118–133. doi: 10.1080/23265507.2017.1354313

Shell Centre (1985). *The language of functions and graphs. An examination module for secondary schools*. Manchester: Joint Matriculation Board by Longman.

Shvarts, A., Chumachemko, D., and Budanov, A. (2014). "The development of the visual perception of the Cartesian coordinate system: An eye tracking study," in *Proceedings of the 38th conference of the international group for the psychology of mathematics education and the 36th conference of the north american chapter of the psychology of mathematics education*, (Canada), 313–320.

Sierpinska, A. (1992). "On understanding the notion of function," in *The concept of function. aspects of epistomology and pedagogy, [maa notes, volume25]*, eds E. Dubinsky and G. Harel (Washington, DC: Mathematical Association of America), 25–58.

Stickler, U., and Shi, L. (2017). Eyetracking methodology in SCMC: A tool for empowering learning and teaching. *ReCALL* 29, 160–177. doi: 10.1017/S0958344017000040

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., and Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educ. Stud. Math.* 104, 147–200. doi: 10.1007/s10649-020-09948-1

Thomaneck, A., Vollstedt, M., and Schindler, M. (2022). *Example of a gaze-overlaid video*. Available online at: https://www.youtube.com/watch?v=RUiMmKU8hWY (accessed November 29, 2022).

Thompson, P. W., and Carlson, M. P. (2017). "Variation, covariation, and functions: Foundational ways of thinking mathematically," in *Compendium for research in mathematics education*, ed. J. Cai (Reston, VA: National Council of Teachers of Mathematics), 421–456.

Tobii Pro AB (2014). *Tobii pro glasses 2 controller software (Version 1.114.20033) [Computer software]*. Danderyd: Tobii Pro AB.

Tobii Pro AB, (2017). *Eye tracker data quality report: Accuracy, precision and detected gaze under optimal conditions—controlled environment*. Danderyd, Sweden: Tobii Pro AB.

Underwood, G., and Everatt, J. (1992). The role of eye movements in reading: Some limitations of the eye-mind assumption. *Adv. Psych. Res.* 88, 111–169. doi: 10.1016/S0166-4115(08)61744-6

van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *Learn. Math.* 25, 2–9.

van Oers, B. (1998). The fallacy of decontextualization. *Mind. Cult. Act.* 5, 135–142. doi: 10.1207/s15327884mca0502_7

van Oers, B. (2001). Contextualization for abstraction. *Cogn. Sci. Q.* 1, 279–305.

Vinner, S., and Dreyfus, T. (1989). Images and definitions for the concept of function. *J. Res. Math. Educ.* 20, 356–366.

Wu, C.-J., and Liu, C.-Y. (2022). Refined use of the eye-mind hypothesis for scientific argumentation using multiple representations. *Instr. Sci.* 50, 551–569. doi: 10.1007/s11251-022-09581-w

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 17 February 2023
DOI 10.3389/feduc.2022.1058150

OPEN ACCESS

EDITED BY
Maike Schindler,
University of Cologne,
Germany

REVIEWED BY
Achim J. Lilienthal,
Örebro University,
Sweden
Jodi Asbell-clarke,
Technical Education Research Centers,
United States
Lonneke Boels,
Utrecht University,
Netherlands

*CORRESPONDENCE
Saskia Schreiter
✉ schreiter@ph-heidelberg.de

SPECIALTY SECTION
This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

RECEIVED 30 September 2022
ACCEPTED 23 December 2022
PUBLISHED 17 February 2023

CITATION
Schreiter S and Vogel M (2023)
Eye-tracking measures as indicators for a local vs. global view of data.
Front. Educ. 7:1058150.
doi: 10.3389/feduc.2022.1058150

Eye-tracking measures as indicators for a local vs. global view of data

Saskia Schreiter* and Markus Vogel

Institute for Mathematics and Computer Science, Heidelberg University of Education, Heidelberg, Germany

Comparing data distributions is a fundamental activity in statistics and a motivating learning opportunity in schools to initiate statistical thinking. Research has shown that many students tend to perceive a data distribution as a collection of individual values rather than as a conceptual entity that has certain features such as center, spread, and shape. These difficulties are reflected in students' tendency to focus on local details of the distribution (so-called *local view of data*) instead of referring to differences between the distributions as a whole (so-called *global view of data*). While many authors refer to school students' conceptions and difficulties related to their *view* of data, there is, to the best of our knowledge, no empirical study that investigated their actual viewing behavior (local vs. global) when comparing distributions. The central assumption of this study is that specific eye-tracking measures constitute indicators for the perceiving and processing of local vs. global distributional features. For this purpose, hypotheses for differences in certain eye-tracking measures (fixation count, saccade amplitude, and saccade direction) between students with a local and global view of data were theoretically derived and empirically investigated using a methodological combination of eye-tracking and stimulated recall interviews. We analyzed data of 25 sixth-grade students who each completed four items on distributional comparisons. The results showed strong positive inter-item correlations for all eye-tracking measures, indicating high internal consistency in participants' gaze behavior across all items. Based on the analysis of the eye-tracking stimulated recall interviews, we split our sample into those students who perceived and processed global features in half or more of the items (global view) and those below that threshold (local view). In line with our theoretically derived hypotheses, students with a global compared to a local view of data had on average significantly fewer fixations, longer saccade amplitudes, and a higher relative number of horizontal saccades. These results suggest that eye-tracking can assist in identifying students' conceptions and difficulties related to a local vs. global view of data. Implications for school practice and further research are discussed.

KEYWORDS

comparing distributions, local and global views of data, statistical thinking, eye-tracking, stimulated recall interviews

1. Introduction

Comparing data distributions is a fundamental activity in statistics (Ben-Zvi, 2004; Burrill and Biehler, 2011). In school, distribution comparisons provide motivating and challenging learning opportunities to initiate statistical thinking already before formal procedures of inferential statistics are known (Konold and Higgins, 2003; Frischemeier, 2019). There is a growing body of research that uses the context of distributional comparison to examine students' conceptions of data distributions and how they relate to their data-based decisions regarding distributional comparison (e.g., Gal et al., 1989; Watson and Moritz, 1999; Ben-Zvi and Arcavi, 2001; Bakker and Gravemeijer, 2004; Ben-Zvi, 2004; Pfannkuch et al., 2004; Frischemeier, 2019). The results of these studies have shown that many students struggle with understanding a data distribution as a whole, as an entity that has many features such as center, spread, and shape. These difficulties are reflected in students' tendency to focus on local details of the distributions (so-called *local view of data*; Bakker and Gravemeijer, 2004; Ben-Zvi, 2004) without attending to the differences between the two distributions as a whole (so-called *global view of data*, Bakker and Gravemeijer, 2004; Ben-Zvi, 2004).

While many authors refer to differences in school students' *view* of data, there is, to the best of our knowledge, no empirical evidence that the perception and processing of local vs. global distributional features can be differentiated based on their specific viewing patterns. Using a combination of eye-tracking and stimulated recall interviews, the present study focuses on students' visual attention and associations with their statistical thinking as they compare distributions and make a data-based decision. The central assumption of this study is that specific eye-tracking measures (fixation count, saccade amplitude, and saccade direction) constitute indicators for the perceiving and processing of local vs. global distributional features. The theoretical associations between those eye-tracking measures and students' statistical thinking are outlined in the following and empirically investigated in the present study.

2. Theoretical background

2.1. Perspectives on data distributions: Local vs. global view

When comparing distributions, several information processing processes are involved, such as perceiving and interpreting features within distributions and putting them in relation between distributions. The distinction between global features that relate to the distribution as a whole (e.g., center, spread, and shape) and local features that refer to single or multiple data points of the distribution (e.g., extreme values and outliers) has been documented and discussed (e.g., Bakker and Gravemeijer, 2004; Ben-Zvi, 2004). These features (local and global) can be approached formally (e.g., calculating the arithmetic mean as a measure of center or the interquartile range as a measure of spread), but also in more exploratory and visual ways (e.g., visually estimating the mean or determining intervals with high density). When comparing two or more distributions, features have to be put in relation between the distributions, adding further relative insights such as regarding *disjoint edge values* (values that are present in one of the distributions and absent from the other), and *common edge values* (the first and last common values of the two distributions; Ben-Zvi, 2004).

An overview of global and local features of distributions that can be considered when comparing data distributions is provided in Table 1. This framework is based on Bakker and Gravemeijer (2004) and has been adapted for the purpose of this study with an explicit focus on visually determinable features of distributions. The structure of the framework can be read upward and downward (*cf.* Bakker and Gravemeijer, 2004): In the upward perspective (local view of data), students typically see the distribution as a collection of individual data points from which they can determine the mean, range, or quartiles, for example. However, this does not necessarily imply that students view these characteristics as measures of center and spread or as representatives of a group (Konold and Pollatsek, 2004). Therefore, it is important that students also develop the downward perspective (global view of data), which is considered essential for statistical data analysis (Ben-Zvi, 2004; Konold et al., 2015). In this perspective, students view data with a notion of distribution as a conceptual entity on its own that has many features such as center, spread/density, and shape. Statistical experts can easily combine the two perspectives (Bakker and Gravemeijer, 2004).

Students' difficulties in understanding a distribution as a whole have been repeatedly documented in research and seem to remain even after instruction in statistics (Ben-Zvi and Arcavi, 2001). For example, students who have already learned to determine formal measures such as the mean and median do not use them when comparing distributions (e.g., Konold et al., 1997; Watson and Moritz, 1999). Even among undergraduate students, difficulties to interpret and compare the variability between two dot plots seem to persist (Lyford, 2017). Instead, students often stick to local additive strategies, also when group sizes are unequal or visual inspection could lead to a straightforward answer (Gal et al., 1989). When comparing data distributions that are represented as dot plots, local strategies repeatedly reported for students include counting and comparing absolute frequencies of dots in certain intervals or calculating and comparing the value of certain intervals of dots (Gal et al., 1989; Watson and Moritz, 1999; Schnell and Büscher, 2015). Likewise, Konold et al. (2015) identified four general perspectives that students from elementary to high school use in working with data. In addition to a global aggregate perspective, these include regarding data as pointers (to the event or context from which the data came), as case values (that provide information on the value of individual cases), and as classifiers (that give information about the frequency of cases that are combined into a new unit).

However, a number of informal strategies that students use to compare groups have been shown and interpreted as steps from a local to a more global view of data. For example, learners use informal terms such as "bumps," "clumps" or "hills" to describe and compare the shape of distributions (e.g., Cobb et al., 2003; Bakker and Gravemeijer, 2004). Konold et al. (2002) and Frischemeier (2019) described how students build ranges in the middle of distributions, so-called "modal clumps", that can be seen as pre-concepts for both center and spread. Bakker and Gravemeijer (2004) demonstrated that students divide distributions into three groups (of low, middle, and high values), which can be interpreted as informal reasoning about density.

While many studies focused on students' statistical reasoning involved in the comparison of data distributions, little is known about the underlying perceptual and attentional processes that

TABLE 1 Perspectives on data distributions: Local vs. global view.

distribution (conceptual entity)		
global features		
center *visually estimated mean, modal value…*	spread/density *observed spread-out-ness, majority, range…*	shape *observed skewness, peaks, tails…*
local features		
individual data points *extreme values, outlier, absolute frequencies of dots in certain intervals, value of certain intervals of dots, disjoint edge values…*		

Global view (arrow from distribution down to local features); *Local view* (arrow from local features up to distribution).

This framework is based on Bakker and Gravemeijer (2004) and has been adapted for the purpose of this study with an explicit focus on visually determinable features of distributions.

guide students in choosing or dismissing features. Theoretically derived associations between specific patterns in students' gaze behavior and their statistical thinking are outlined in the following.

2.2. Eye-tracking for investigating students' mathematical thinking

Numerous studies in mathematics education research have shown that eye-tracking has the potential to provide new insights into students' mathematical thinking and learning (for an overview, see Lilienthal and Schindler, 2019; Strohmaier et al., 2020). However, existing research does not cover all mathematical topics in the same depth. The field of statistics, for example, was identified as a domain with rarely any eye-tracking studies (Strohmaier et al., 2020), although aspects of visualization and mental representations play a vital role in this domain. There are just a few very recent studies that used eye-tracking technology to study students' strategies and difficulties when interpreting and comparing statistical graphs such as histograms (e.g., Boels et al., 2019; Lyford and Boels, 2022). Eye-tracking is a suitable method to obtain information about visual attention and cognitive processing while students are solving problems, especially when visual strategies are involved (e.g., Andrà et al., 2009; Klein et al., 2018; Malone et al., 2020). A major advantage of the eye-tracking method is that students' solution processes can be observed without interrupting them (e.g., Inglis and Alcock, 2012; Obersteiner and Tumpek, 2016). Furthermore, cognitive processes in mathematical thinking are often complex and may occur unconsciously. Nevertheless, these processes are reflected in students' eye movements (e.g., Ott et al., 2018; Schindler and Lilienthal, 2019).

The most used eye-tracking measures are derived from fixations and saccades. A fixation is a period of time during which the eye remains relatively still on a visual stimulus and information can be absorbed (Holmqvist and Andersson, 2017). Saccades are periods where the eye moves very fast, lasting less than 100 ms (Holmqvist and Andersson, 2017). The amount of attention to specific objects is often measured by the *number of fixations* on these objects (e.g., Andrá et al., 2015; Schindler and Lilienthal, 2019). Furthermore, in their systematic review on eye-tracking in mathematics education research Strohmaier et al. (2020, p. 17) found that *saccade amplitude* is "used as an indicator for local (short saccades) compared to global (long saccades) strategies in information retrieval (Inglis and Alcock, 2012; Stolińska et al., 2014; Klein et al., 2018) and information integration (Godau et al., 2014)." *Saccade direction* is interpreted as indicative of search strategy (Poole and Ball, 2005) or the strategy of how a diagram is read (Klein et al., 2018). In addition, Khalil (2005) found that experts in comparison to novice data analyzers show more horizontal and less vertical search movements when visually inspecting and comparing graphs. This was also related to the observation that experts tend to use more global comparison methods whereas novices tend to use more local methods to compare data distributions.

It is important to consider that while eye-tracking data is rich in information, it is also complex and not always unambiguously interpretable (e.g., Schindler and Lilienthal, 2019; Strohmaier et al., 2020). Therefore, it is recommended for eye-tracking research to formulate clear hypotheses about the expected eye movement patterns based on theory and previous studies, rather than taking an exploratory approach (Orquin and Holmqvist, 2017). In addition, a methodological triangulation with other research methods, such as stimulated recall interviews, is recommended (e.g., Wyss et al., 2021). In an eye-tracking stimulated recall interview, participants are asked to retrospectively describe their own thoughts based on a video sequence of their eye movements (Hyrskykari et al., 2008). Visualizing eye movements serves as a memory aid to recall own actions and thoughts and has proven to be an effective method to stimulate reflection of internal cognitive processes (e.g., van Gog et al., 2005; Schreiter et al., 2022).

2.3. Eye-tracking measures as indicator for a local vs. global view of data

Students with a local compared to a global view of data allocate their attention on individual data points instead of viewing the distribution as a whole (e.g., Bakker and Gravemeijer, 2004; Ben-Zvi, 2004). Building on the above-illustrated theoretical associations between eye-tracking measures and cognitive processes (*cf.* 2.2), the aim of the present study is to explore students' visual attention and related statistical thinking when comparing data distributions. Table 2 provides an overview of the above-introduced eye-tracking measures, their definitions (according to Holmqvist and Andersson, 2017), their theoretically derived interpretations (*cf.* 2.2), and their relevance for the research interest of this study. As the distributions presented in the study items consist of a high number of individual data points, students with a local compared to a global view of data should show a higher number of fixations, indicating that more attention is paid to the individual data points that make up the distribution (e.g., Schindler and Lilienthal, 2019). Saccade amplitude is interpreted as an indicator of local (short saccades) compared to global (long saccades) strategies in information retrieval (Strohmaier et al., 2020). Saccade direction is used as indicative of search strategy (Poole and Ball, 2005) or the strategy of how a diagram is read (Klein et al., 2018). To perceive global features such as center, spread, and shape, the distribution needs to be viewed as a whole. For the distributions shown in the items of this study (see Figure 1), this should result in longer saccade amplitudes and more saccades in the horizontal direction.

3. The present study

The present study focused on students' visual attention and associations with their statistical thinking (operationalized in this study as the perception and processing of local vs. global distributional features). Four items were constructed that involve a comparison of data distributions and require a data-based decision. We chose a methodological combination of eye-tracking and stimulated recall interviews—an approach that was shown to be effective to give insights into students' visual attention and related cognitive processes. The following research questions were addressed:

> *RQ1*: Which distributional features (local vs. global) do students perceive and process?
>
> *RQ2*: Can the perception and processing of local vs. global distributional features (local vs. global view of data) be distinguished based on specific eye movement measurements?

Based on the theoretical considerations presented in chapter 2.3, we hypothesized that students with a global view compared to a local view of data have fewer fixations (H1), longer saccade amplitudes (H2), and a higher relative number of horizontal saccades (H3).

Moreover, we assume that it is more time consuming to focus attention on many individual data points compared to the distribution as a whole. Therefore, we expect that a local compared to a global view of data is characterized by longer total viewing times (i.e., the average time needed by the students to complete a task) (H4).

The following sections provide details on the design of the study setup and the quantitative and qualitative analysis used to investigate our hypotheses.

4. Methods

4.1. Sample

The sample consisted of $N = 25$ 6th-grade students (56% female) from two German secondary schools. On average, participating students were 11.6 years old ($SD = 0.57$). The schools were of type *Gymnasium*, the highest track of secondary education in the German school system. According to their curriculum, students had been formally introduced to determine specific local distributional features (e.g., maximum, minimum) and global features of center (e.g., arithmetic mean), but not yet to determine measures of spread that are typically not introduced until grades 7/8. Students were recruited through their mathematics teachers, who agreed to participate in the study with their class. Participation was voluntary, all students had normal or correct-to-normal vision.

4.2. Material

Four items on distribution comparisons were developed for the purpose of this study. An example item is presented in Figure 1

TABLE 2 Eye-tracking measures, definitions, interpretations, and potential relevance for distribution comparisons.

Eye-tracking measure	Definition	Interpretation	Relevance for distribution comparisons
Number of fixations	Frequency, how often an object/area is fixed	Amount of attention on this object/area	Higher/lower number of fixations as an indicator of a local/global view of data where attention is focused on many individual data points/the distribution as a whole
Saccade amplitude	The distance (in pixels) travelled by a saccade from its onset to offset	Indicator for local (short saccades) compared to global (long saccades) strategies in information retrieval and integration	Indicator for local (short saccades) compared to global (long saccades) strategies in comparing distributions
Saccade direction	Absolute angle (in degrees) of a saccade measured to the horizontal	Indicator for search strategy/reading strategy of a diagram	Horizontal/vertical saccade directions as indicator for perception and processing of global/local features of the distribution

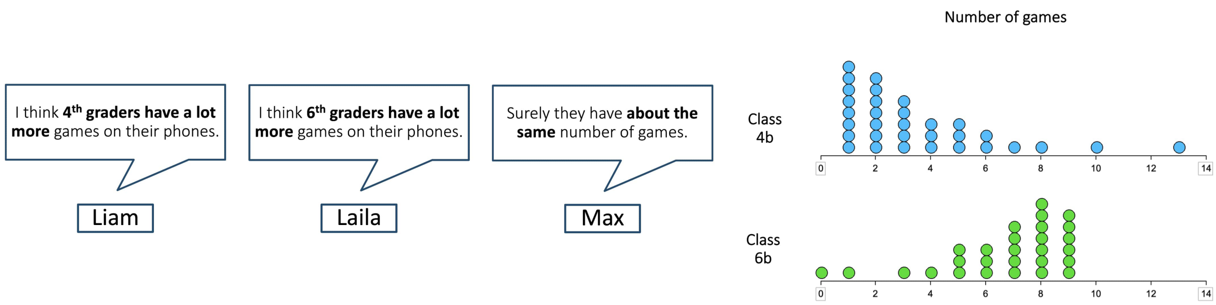


FIGURE 1
Sample item. Participants (N=25) were shown three assumptions of children (left) before and after they analyzed and compared the distributions (right). Eye-tracking data were collected only while the distributions were presented.

(translated from the original German into English). All items include authentic comparison situations: data from a hypothetical survey regarding the topic "games on smartphone" are presented and compared between different groups of school students. In each item, participants are first presented with three assumptions of children and then they are shown data from the survey (*cf.* Figure 1). All tasks include an explicit request to draw a conclusion from the data presented in the task. Participants are asked to compare the samples and decide to which of the previously presented assumptions (in the example item: Liam, Laila, or Max, *cf.* Figure 1) the data fit. Features in which the distributions differ (such as center, spread, or shape) were varied between the items. In addition, the sample sizes (equal/unequal) were varied, so that one item with very different sample sizes was used, two items with slightly different sample sizes, and one item with equal sample sizes. Students were explicitly given the information of sample sizes before each item. This variation was chosen to test whether students might switch flexibly between local and global strategies depending on certain characteristics of the distribution comparison. For example, while comparing absolute frequencies of dots in certain intervals or comparing the value of certain intervals of dots are valid local strategies to compare samples of equal sizes, it is incorrect to do so when sample sizes are unequal. A pilot study was conducted with N=20 grade 6 students to assess the comprehensibility of the items and study procedure.

4.3. Procedure

The study was conducted in the children's schools. An environment to which the students were used should help reduce anxiety and nervousness. Participants received instructions and the four items were presented in randomized order on a 24-inch computer screen (Fujitsu B24T-7 LED, 1920×1,080 pixels). An example item was used to explain the study procedure to all participants in advance. The average distance from the participant to the monitor was around 60 cm. No chin rest was used as the system allows high quality of measurement accuracy even with smaller head movement. While participants worked on the tasks (which took an average 75.36 s per item, SD =73.03), their eye movements were recorded with a monitor-based eye-tracker (Tobii pro fusion). The eye-tracker captured gaze data with a sampling frequency of up to 120 Hz and an average accuracy of 0.74° (SD =0.47°). Before each task, a 9-point calibration was performed to achieve optimal recording of the eye-tracking data (Holmqvist and Andersson, 2017).

An eye-tracking stimulated recall interview was conducted directly following each task. Conducting eye-tracking stimulated recall interviews is an effective research method to investigate students' cognitive processes by asking them to retrospectively describe their own thoughts and actions as precisely as possible based on a video of their own gaze movements (Schindler and Lilienthal, 2019). The core idea of eye-tracking stimulated recall interviews is to provide reflection-aiding stimuli (van Gog et al., 2005). For the interviews, we used the videos recorded by *Tobii pro lab* software (Tobii Pro, 2014). Eye movements were displayed by a red dot and lines connecting gaze points. During the interview, the interviewer showed a positive and interested attitude in the children's descriptions without judging them. Both the interviewer and student could pause in the video, for example, if a participant wanted to describe in more detail the thoughts associated with a particular scene. For the recording of the interviews, Open Broadcaster Software (OBS)[1] was used, which recorded image screen contents including sound, so that the videos of the gaze movements with the corresponding comments of the participants were available for the analysis.

4.4. Data and statistical analyses

4.4.1. Eye-tracking data

Analysis of the eye-tracking data was conducted with *Tobii pro lab* software. In each item, two areas of interest (AOIs) were created covering the two data distributions.

An I-VT algorithm (Salvucci and Goldberg, 2000) was used to detect fixations and saccades within the AOIs. Using the Tobii I-VT fixation filter (Tobii Pro AB, 2014), eye movements were classified either as part of a fixation if the velocity is below the threshold of

1 https://obsproject.com

30°/s, or as part of a saccade if the velocity is equal or higher than this threshold. To determine the saccadic measures (saccade amplitude and saccade direction), raw gaze data were analyzed. Only saccades that have the immediately preceding and consecutive fixations within the same AOI were considered for the analysis. Saccades that start in one AOI and end in the other AOI are called transitions (Holmqvist and Andersson, 2017), and provide information about how distributional features (local or global) are put in relation between the two distributions. These saccades were not considered, as they are not relevant to the research interest of this study. Saccade direction was defined as the absolute angle of a saccade (in degrees) measured to the horizontal and calculated based on the coordinates of the immediately preceding and consecutive fixations using basic trigonometry (*cf.* Holmqvist and Andersson, 2017, p. 440f). Saccades were classified as horizontal if this angle was between 0° and |44|° and classified as vertical if it was between |45|° and |90|°.

One recording of one participant had to be excluded due to data loss, as the eye-tracker lost track of the participants' eye.

4.4.2. Eye-tracking stimulated recall interview

Conducting eye-tracking stimulated recall interviews is an effective method to get insights into students' mathematical thinking (*cf.* 2.2). In this study, this methodological approach was used to gain insight into the statistical thinking of students regarding the perception and processing of local and global distributional features. A mixed-methods approach was used to analyze the data. The videos of the interviews contained the gaze movements with the corresponding comments of the participants and were coded both deductively and inductively using qualitative content analysis (Rädiker and Kuckartz, 2019). Table 3 shows sample screenshots of gaze plots, related excerpts from students' comments, and assigned categories. As the focus of the research presented here was on students' statistical thinking while comparing data distributions, the coding procedure only referred to the process until the distribution comparison decision was made. The decision of the participants was not included in the data analysis presented here.

In total, 98 videos were coded. Two videos were excluded due to technical problems with screen recording.

The transcripts were coded by two raters with high interrater reliability (Cohen's kappa = 0.87). The assigned codes were integrated into a quantitative data set to examine differences in viewing behavior (fixation count, saccade amplitude, and saccade direction) between students who perceived and processed predominantly local features compared to students who perceived and processed predominantly global features.

5. Results

5.1. Visual attention when comparing data distributions

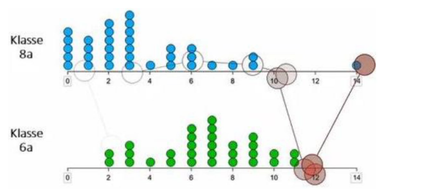
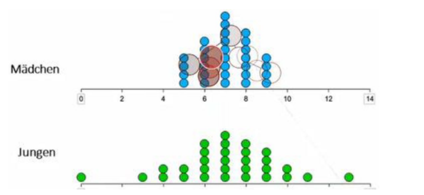
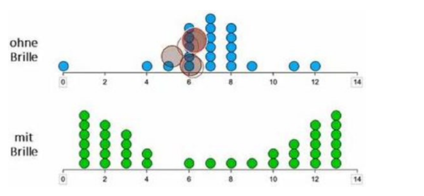
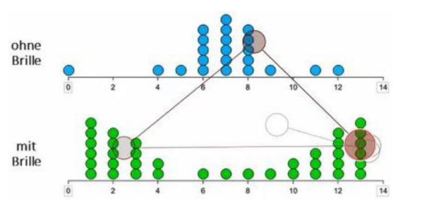
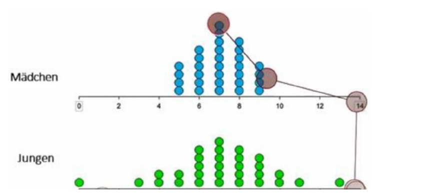
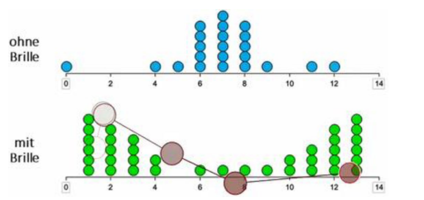
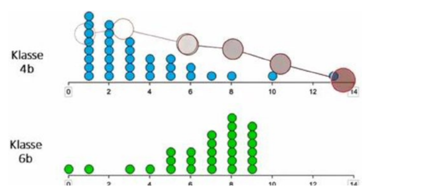
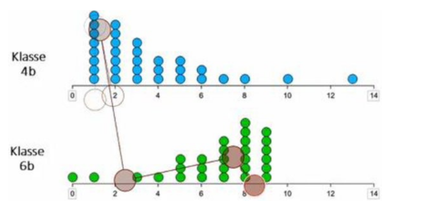
In this study, specific eye-tracking measures were captured giving insight into students' visual attention when comparing data distributions. Table 4 displays mean scores and standard deviations for the number of fixations, saccade amplitudes, and relative number of horizontal saccades (within AOIs), as well as for the total viewing time (i.e., the average time needed by the students to complete a task). High standard deviations indicate high interindividual differences in gaze behavior between participants. To measure to what degree the viewing behavior of participants is interrelated across all four items used in this study, inter-item Pearson's correlations were calculated. The results show strong positive inter-item correlations for all eye-tracking measures, indicating high internal consistency in participants' gaze behavior across all items (Table 4). This implies that although the participating students show very different gaze behavior among themselves, the same child remains relatively constant across items. Thus, the collected eye-tracking measures appear to be comparable across the four items considered in this study.

5.2. Statistical thinking when comparing data distributions

To gain information on students' statistical thinking with regard to the perception and processing of global and local distributional features, we analyzed the eye-tracking stimulated recall interviews in a qualitative manner. The results showed that out of the 25 participants, eight students did not consider global features in any of the four items, two students considered at least one global feature in one of the items, three students in two of the items, three students in three of the items, and nine students in all four items. Thus, the vast majority of the students (68%) did either not consider global features in any of the items or in all four items. This showed that most students stayed relatively constant in their comparison strategy across all items. The students who did not consider any global characteristics across all items are to be classified as problematic in this context. These students remained with local strategies (e.g., comparing absolute frequencies of dots in certain intervals or comparing the value of certain intervals of dots), even if sample sizes are unequal, which is an incorrect strategy in these cases.

Regarding global features, students' utterances were assigned to the three categories center, spread/density, and shape. The category *spread/density* was often assigned as students divided the distributions into three groups (of low, middle, and high values) or identified and compared areas with particularly many dots (so-called modal clumps). Sometimes, students also referred to range and compared how "spread out" or how "close together" the data points of the distribution are. The category *center* was mainly assigned to students who identified and compared the modal values of the two distributions. Only once the category was assigned to a student who visually estimated and compared the means of both distributions. Regarding shape, students' utterances were very different, comparing the shape of the distributions for example to stairs that go up and down or to a deckchair (first down, then up again). Sometimes students also chose mathematical terms such as "symmetrical" or "triangle shaped" to describe the distributions' shapes.

TABLE 3 Analysis of the eye-tracking stimulated recall interviews.

Screenshot of gaze plot	Related excerpt from the students' comments	Assigned categories (sub-categories)
	*"Here I looked who has the most games and saw that one person has 14 games and the others (from class 6a) have less…"*	Local *(extreme values: maximum)*
	*"With the girls, there are four that have five games, six that have six games, nine that have seven games..."*	Local (*absolute frequencies of dots in certain intervals*)
	*"I first calculated four plus five, that's nine and here I counted five people have six games, then I calculated five times six, that's 30 and I added the nine from before and then…"*	Local (*value of certain intervals of dots*)
	*"I first looked at those without glasses and then I saw that they have relatively many in the middle but on the outside, that is, with more or less games, they have only a few people and then I looked at those with glasses and then I saw that they have fewer people in the middle but many with less or with a lot of games..."*	Global *(spread: division in three groups, majority)*
	*"Here I noticed that they (the girls) were very close together, and not so widely distributed and below (with the boys) it went rather in the width and above (with the girls) in the height"*	Global *(spread: observed spread-out-ness)*
	*"With the others (with glasses) it was more in the form of a deck chair, it first goes down and then up again"*	Global *(shape)*
	*"Here I noticed that it is staircase-shaped (...) then I compared the height of the towers…"*	Global *(shape)*
	*"Then I looked where is the most, that is at the 1 (for class 4b), then I looked at the bottom where is the most for class 6, that is at the 8"*	Global *(center: modal value)*

Sample screenshots of gaze plots, related excerpts from students' comments, and assigned categories (local vs. global feature) and sub-categories.

TABLE 4 Means, standard deviations, and inter-item Pearson's correlations for eye-tracking measures.

ET measure		*M* (*SD*)	Item 1	Item 2	Item 3	Item 4
Number of fixations	Item 1	152.92 (140.60)	–			
	Item 2	133.96 (116.27)	0.70***	–		
	Item 3	157.24 (130.22)	0.87***	0.86***	–	
	Item 4	161.00 (141.03)	0.82***	0.90***	0.88***	–
Saccade amplitude (pixels)	Item 1	103.67 (30.19)	–			
	Item 2	107.50 (39.47)	0.71***	–		
	Item 3	97.40 (35.56)	0.80***	0.78***	–	
	Item 4	116.06 (48.90)	0.77***	0.77***	0.77***	–
Relative number of saccades in horizontal direction	Item 1	0.61 (0.13)	–			
	Item 2	0.59 (0.12)	0.52**	–		
	Item 3	0.53 (0.13)	0.57**	0.38*	–	
	Item 4	0.59 (0.14)	0.64***	0.39*	0.47*	–
Total viewing time (s)	Item 1	75.87 (75.78)	–			
	Item 2	67.71 (70.86)	0.73***	–		
	Item 3	77.85 (76.12)	0.89***	0.87***	–	
	Item 4	80.02 (87.16)	0.84***	0.84***	0.92***	–

$N = 25$ for Item 1, Item 2, Item 3. $N = 24$ for Item 4.
***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

TABLE 5 Descriptive statistics and *t*-test results for eye-tracking measures of students with a local vs. global view of data.

ET measure	Local view		Global view		*t*(23)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Number of fixations	232.63	144.66	95.58	67.64	2.80	0.016	1.21
Saccade amplitude (pix.)	78.12	17.43	124.68	30.85	−4.81	<0.001	1.86
Relative number of saccades in horizontal direction	0.51	0.09	0.63	0.08	−3.28	0.003	1.41
Total viewing time (s)	130.26	82.85	38.77	33.81	3.85	<0.001	1.45

Descriptive statistics and results of two-tailed *t*-tests are shown for students with a global ($n = 15$) and local view of data ($n = 10$) across all items.

5.3. Eye-tracking measures as indicators for a local vs. global view of data

The central assumption of this study was that the perception and processing of local vs. global features can be distinguished using specific eye-tracking measures. To test this assumption, we split our sample into those students who perceived and processed at least one global feature in at least half of the items (from now on referred to as students with a *global view*) and those students below that threshold (from now on referred to as students with a *local view*). Subsequently, two-tailed *t*-tests were calculated to test whether students with a local and global view differed with respect to the collected eye-tracking measures. The results show significant group differences with high effect sizes for all collected eye-tracking measures (Table 5).

The empirical data confirmed all of our theoretically derived hypotheses: students with a global compared to a local view of data showed on average significantly fewer fixations (H1), longer saccade amplitudes (H2), and a higher relative number of horizontal saccades (H3). These group differences consistently showed high effect sizes. In line with our expectations, significant group differences also emerged at the total viewing time. Students with a global view of data needed less than half the time to draw a data-based decision regarding the distribution comparison compared to students with a local view of data (*cf.* Table 5). However, there was a strong positive correlation between total viewing time and the number of fixations ($r = 0.905$, $p < 0.001$), which is why these measurements cannot be considered independent of each other.

6. Discussion

This study investigated students' visual attention and statistical thinking while comparing data distributions. The central assumption was that specific eye-tracking measures constitute indicators for the perception and processing of local and global distributional features (local vs. global view of data). In accordance with recommendations for the methodological approach of eye-tracking studies (e.g., Orquin and Holmqvist, 2017, see chapter 2.2), we theoretically derived hypotheses for differences in certain eye-tracking measures (fixation count, saccade amplitude, and saccade direction) between students with a local and global view of data and empirically investigated these using a methodological combination of eye-tracking and stimulated recall interviews.

With regard to the first research question, we analyzed which distributional features (local vs. global) students perceived and processed. The findings of the eye-tracking stimulated recall interviews revealed that most of the students did either not consider global features in any of the items or in all four items. Thus, students showed a certain consistency in their statistical thinking across all items. An essential characteristic of statistical data analysis is that it is mainly about describing global features of data distributions (e.g., Bakker and Gravemeijer, 2004; Ben-Zvi, 2004). Against this background, students who did not consider any or only rarely considered global characteristics across all items are to be classified as problematic. These students remained with local strategies (e.g., comparing absolute frequencies of dots in certain intervals), even if sample sizes are unequal, which is an incorrect strategy in these cases. These findings are in line with existing research that described students' difficulties in understanding a distribution as a whole which seem to persist even after instruction in statistics (e.g., Konold et al., 1997; Watson and Moritz, 1999; Ben-Zvi and Arcavi, 2001). In their study, Gal et al. (1989, p. 6) found that "many students blindly added even when groups were of unequal sizes (…) even when a visual inspection of the data could lead to a straightforward decision." Similar observations can be reported for many students that participated in this study. Before each item, students were explicitly given the information of sample sizes (which was unequal in three of four items). Nevertheless, students who chose local strategies in one item mostly sticked to them with the other items.

However, several students also used global features, such as center, spread/density, or shape to compare distributions. For example, students divided the distributions into three groups (of low, middle, and high values) or identified and compared areas with particularly many data points (so-called modal clumps). Similar strategies were observed in other studies (e.g., Konold et al., 2002; Bakker and Gravemeijer, 2004; Frischemeier, 2019) and interpreted as informal reasoning about spread and density. Sometimes, students also referred to range and used informal terms such as "spread out" or "close together" to describe the distributions. Regarding features of the center, students mostly compared the modal values of the two distributions, while a visual estimation of the arithmetic mean was hardly performed. Moreover, students also used informal terms such as "stair-case shaped" or more formal terms such as "symmetric" and "triangle-shaped" to describe and compare the shape of distributions. Similar attempts to describe a distribution's shape were also observed in previous studies (e.g., Cobb et al., 2003; Bakker and Gravemeijer, 2004) and interpreted as steps from a local to a more global view of data.

To address our second research question, we investigated whether the perception and processing of local vs. global distributional features (local vs. global view of data) could be distinguished based on specific eye movement measures. Eye-tracking has proven to be an effective method to obtain information about students' visual attention and cognitive processing while solving problems, especially when visual strategies are involved (e.g., Klein et al., 2018; Schindler and Lilienthal, 2019; Malone et al., 2020). We first wanted to check if the collected eye-tracking measures are comparable across the four items considered in this study. The analysis of the collected eye-tracking measures indicated high interindividual differences between the participants. These were evident for all four items. At the same time, results showed strong positive inter-item correlations for all eye-tracking measures, indicating high internal consistency in participants' gaze behavior across all items. Consequently, although the participating students show very different gaze behavior among themselves, one and the same child remains relatively constant across items. Thus, the collected eye-tracking measures appear to be comparable across the four items.

Based on the analysis of the eye-tracking stimulated recall interviews, the sample was split into those students who perceived and processed global features in half or more of the items (global view) and those below that threshold (local view). In line with our theoretically derived hypotheses, students with a global compared to a local view of data on average had significantly fewer fixations (H1), longer saccade amplitudes (H2), and a higher relative number of horizontal saccades (H3). All group differences consistently showed high effect sizes. These results suggest that eye-tracking data can assist in identifying students' conceptions and difficulties related to a local vs. global view of data. While many authors refer to school students' conceptions and difficulties related to their *view* of data, this is, to the best of our knowledge, the first empirical study that investigated their actual viewing behavior in relation to their local vs. global strategies when comparing distributions. Understanding which features students attend their visual attention on and what is happening in students' minds while they are visually focusing these features may provide further insights into how task design and instruction should be structured to guide students from a local to a global view on data, which is considered an important goal of statistics education (Ben-Zvi and Arcavi, 2001). Furthermore, the results of this study can provide an initial basis for the potential of eye-tracking as a diagnostic tool for detecting students' conceptions and difficulties in distributional comparison.

As expected, significant group differences also emerged for total viewing time. Students with a global view of data took less than half the time to make a data-based decision regarding the distributional comparison than students with a local view of data. However, as the total viewing time is strongly correlated with the number of fixations, these measurements cannot be considered independent.

6.1. Limitations and implications for further research

We would like to emphasize that the findings of this study should be considered in the light of some limitations.

The results and statistics of the study must be interpreted in light of the relatively small sample size. Potential influencing factors (e.g., age and topic-specific pre-knowledge) on students' gaze behavior and on their performance in perceiving and processing local and global distributional features were not considered and should be investigated in future studies. It would also be interesting to study whether the presented eye-tracking measures can be applied to other data visualizations as indicators of a local and global view of data.

In addition, we only differentiated between students that perceived at least one global feature in half or more of the items (global view) and those students below that threshold (local view). However, also within these groups, students showed large differences. For example, within the group of global viewers, the performance of a student who only compared the modal values of the two distributions can be classified as less high than that of a student who considered several global features of center, spread/density, and shape and related them to each other. Future research should therefore examine performance differences within the groups of local and global viewers in more detail. This could include the number of global/local features considered, their statistical nature (center, spread/density, shape), and a distinction as to whether the perceived local/global features are put in relation or considered in isolation to make a decision regarding the distributional comparison.

6.2. Conclusion and future directions

The present study should be seen as a first step toward enhancing our understanding of students' visual attention and associated statistical thinking when comparing data distributions. The collected empirical data supported our theoretically derived hypotheses, showing that students with a global compared to a local view of data had significantly fewer fixations, longer saccade amplitudes, and a higher relative number of horizontal saccades. These results suggest that eye-tracking can assist in identifying students' conceptions and difficulties related to a local vs. global view of data. Future research is necessary to study performance differences in students' statistical thinking in more detail, including potential influencing factors on the part of students. Furthermore, the results of this study could serve as a starting point for future research that investigates the potential of eye-tracking as a diagnostic tool that can be used in teacher training or in school practice to detect and learn about students' conceptions and difficulties in distributional comparison.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

Both authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Funding

This study was carried out as part of the graduate school "DiaKom" funded by the Ministry of Science, Research, and the Arts in Baden-Wuerttemberg, Germany.

Acknowledgments

We would like to thank all students and their teachers for participating in and supporting this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Andrà, C., Arzarello, F., Ferrara, F., Holmqvist, K., Lindström, P., Robutti, O., et al. (2009). "How students read Mathematical Representations: An eye tracking study." in *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education (Vol. 2). Aristotle University of Thessaloniki & University of Macedonia, Thessaloniki, Greece.* eds M. Tzekaki, M. Kaldrimidou, & C. Sakonidis, 49–56

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., and Sabena, C. (2015). Reading mathematics representations: an eye-tracking study. *Int. J. Sci. Math. Educ.* 13, 237–259. doi: 10.1007/s10763-013-9484-y

Bakker, A., and Gravemeijer, K. P. E. (2004). "Learning to reason about distribution" in *The challenge of developing statistical literacy, reasoning and thinking.*

eds. D. Ben-Zvi and J. Garfield (Netherlands: Springer), 147–168. doi: 10.1007/1-4020-2278-6_7

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Stat. Educ. Res. J.* 3, 42–63. Available at: https://www.academia.edu/5736781/Reasoning_about_variability_in_comparing_distributions

Ben-Zvi, D., and Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educ. Stud. Math.* 45, 35–65. doi: 10.1023/A:1013809201228

Boels, L., Bakker, A., and Drijvers, P. (2019). "Eye tracking secondary school students' strategies when interpreting statistical graphs" in *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education (Vol. 2). Pretoria, South Africa.* eds. M. Graven, H. Venkat, A. Essien and P. Vale, 113-120.

Burrill, G., and Biehler, R. (2011). "Fundamental statistical ideas in the school curriculum and in training teachers" in *Teaching statistics in school mathematics – Challenges for teaching and teacher education.* eds. C. Batanero, G. Burrill and C. Reading (Dordrecht: Springer), 57–69.

Cobb, P., McClain, K., and Gravemeijer, K. (2003). Learning about statistical Covariation. *Cogn. Instr.* 21, 1–78. doi: 10.1207/S1532690XCI2101_1

Frischemeier, D. (2019). Primary school students' reasoning when comparing groups using modal clumps, medians, and hatplots. *Math. Educ. Res. J.* 31, 485–505. doi: 10.1007/s13394-019-00261-6

Gal, I., Rothschild, K., and Wagner, D. A. (1989). "Which group is better? The development of statistical reasoning in elementary school children" in *Meeting of the Society for Research in Child Development* (Kansas City, MO) https://eric.ed.gov/?id=ED315270

Godau, C., Wirth, M., Hansen, S., Haider, H., and Gaschler, R. (2014). From marbles to numbers—estimation influences looking patterns on arithmetic problems. *Psychology* 05, 127–133. doi: 10.4236/psych.2014.52020

Holmqvist, K., and Andersson, R. (2017). *Eye Tracking: A Comprehensive Guide to Methods, Paradigms, and Measures. 2nd Edn.* Lund, Sweden: Lund Eye-Tracking Research Institute.

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K.-J., and Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *J. Eye Mov. Res.* 2, 1–18. doi: 10.16910/jemr.2.4.5

Inglis, M., and Alcock, L. (2012). Expert and novice approaches to Reading mathematical proofs. *JRME* 43, 358–390. doi: 10.5951/jresematheduc.43.4.0358

Khalil, K. A. I. (2005). *Expert-novice Differences: Visual and Verbal Responses in a Two-Group Comparison Task [Master's thesis].* University of Massachusetts: Amherst, MA, USA. Available at: https://scholarworks.umass.edu/theses/2428

Klein, P., Viiri, J., Mozaffari, S., Dengel, A., and Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: how do students look at vector field plots? *Phys. Rev. Phys. Educ. Res.* 14:10116. doi: 10.1103/PhysRevPhysEducRes.14.010116

Konold, C., and Higgins, T. L. (2003). "Reasoning about data" in *A research companion to principles and standards for school mathematics*. eds. J. Kilpatrick, W. G. Martin and D. Schifter (Reston, VA: National Council of Teachers of Mathematics), 193–215.

Konold, C., Higgins, T., Russell, S. J., and Khalil, K. (2015). Data seen through different lenses. *Educ. Stud. Math.* 88, 305–325. doi: 10.1007/s10649-013-9529-8

Konold, C., and Pollatsek, A. (2004). "Conceptualizing an average as a stable feature in a Noisy process" in *The challenge of developing statistical literacy, reasoning and thinking*. eds. D. Ben-Zvi and J. Garfield (Netherlands: Springer), 169–199. doi: 10.1007/1-4020-2278-6_8

Konold, C., Pollatsek, A., Well, A., and Gagnon, A. (1997). "Students analyzing data: research of critical barriers" in *Research on the role of Technology in Teaching and Learning Statistics: Proceedings of the 1996 IASE round table conference University of Granada, Spain.* eds. J. Garfield and G. Burrill (Voorburg, The Netherlands: International Statistical Institute), 151–167.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002). "Students use of modal clumps to summarize data." in *Developing a Statistically Literate Society: Proceedings of the Sixth International Conference on Teaching of Statistics, Cape Town, South Africa, 7-12 July 2002.* eds. B. Philips (Voorburg, Netherlands: International Assoc. for Statistical Teaching), 1–6.

Lilienthal, A. J., and Schindler, M. (2019). "Eye tracking research in mathematics education: A PME literature review." in *Proceedings of 43rd Conference of the International Group for the Psychology of Mathematics Education (Vol 4). Pretoria, South Afrika.* eds. M. Graven, H. Venkat, A. Essien, and P. Vale, 62.

Lyford, A. J. (2017). Investigating undergraduate student understanding of graphical displays of quantitative data through machine learning algorithms. Doctoral dissertation, University of Georgia. Available at: https://getd.libs.uga.edu/pdfs/lyford_alexander_j_201705_phd.pdf

Lyford, A., and Boels, L. (2022). "Using machine learning to understand students' gaze patterns on graphing tasks." in *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* eds. S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (International Association for Statistical Education), 1–6.

Malone, S., Altmeyer, K., Vogel, M., and Brünken, R. (2020). Homogeneous and heterogeneous multiple representations in equation-solving problems: an eye-tracking study. *J. Comp. Assist. Learn.* 781–798. doi: 10.1111/jcal.12426

Obersteiner, A., and Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM* 48, 255–266. doi: 10.1007/s11858-015-0742-z

Orquin, J. L., and Holmqvist, K. (2017). Threats to the validity of eye-movement research in psychology. *Behav. Res. Methods* 50, 1645–1656. doi: 10.3758/s13428-017-0998-z

Ott, N., Brünken, R., Vogel, M., and Malone, S. (2018). Multiple symbolic representations: the combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learn. Instr.* 58, 88–105. doi: 10.1016/j.learninstruc.2018.04.010

Pfannkuch, M., Budgett, S., Parsonage, R., and Horring, J. (eds.) (2004). Comparison of data plots: Building a pedagogical framework. Paper presented at ICME-10, TSG11: Research and development in the teaching and learning of probability and statistics. Online available at: https://iase-web.org/Conference_Proceedings.php?p=ICME_10_2004.

Poole, A., and Ball, L. J. (2005). "Eye tracking in human-computer interaction and usability research: current status and future prospects" in *Encyclopedia of human-computer interaction*. ed. C. Gahoui (Hershey, Pa: IGI Global), 211–219.

Rädiker, S., and Kuckartz, U. (2019). Analyse qualitativer Daten mit MAXQDA. *Springer Fachmedien Wiesbaden*. doi: 10.1007/978-3-658-22095-2

Salvucci, D. D., and Goldberg, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols" in *Proceedings of the 2000 symposium on eye tracking research & applications* ETRA '00 ed. A. T. Duchowski (New York, NY: ACM), 71–78.

Schindler, M., and Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educ. Stud. Math.* 101, 123–139. doi: 10.1007/s10649-019-9878-z

Schnell, S., and Büscher, C. (2015). "Individual concepts of students comparing distributions" in *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic.* eds. K. Krainer and N. Vondrová, 754–760.

Schreiter, S., Vogel, M., Rehm, M., and Dörfler, T. (2022). Die Rolle des Wissens angehender Mathematiklehrkräfte beim Diagnostizieren schwierigkeitsgenerierender Aufgabenmerkmale. Erkenntnisse aus Eye-Tracking Stimulated Recall Interviews. *J. Math.-Didakt.* 43, 101–133. doi: 10.1007/s13138-022-00203-z

Stolińska, A., Andrzejewska, M., Błasiak, W., Pęczkowski, P., Rosiek, R., Rożek, B., et al. (2014). "Analysis of saccadic eye movements of experts and novices when solving text tasks" in *New Technologies in Science Education*. eds. M. Nodzyńska, P. Cieśla, and K. Różowicz (Cracow: Pedagogical University of Cracow), 21–29.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., and Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: a systematic literature review. *Educ. Stud. Math.* 104, 147–200. doi: 10.1007/s10649-020-09948-1

Tobii Pro AB. (2014). *Tobii Pro lab (version 1.162.32461 (x64)) [computer software]*. Danderyd, Sweden: Tobii Pro AB.

van Gog, T., Paas, F., van Merriënboer, J. J. G., and Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting vs. concurrent and retrospective reporting. *J. Exp. Psychol. Appl.* 11, 237–244. doi: 10.1037/1076-898X.11.4.237

Watson, J. M., and Moritz, J. B. (1999). The beginning of statistical inference: comparing two data sets. *Educ. Stud. Math.* 37, 145–168. Available at: http://www.jstor.org/stable/3483313

Wyss, C., Rosenberger, K., and Bührer, W. (2021). Student teachers' and teacher educators' professional vision: findings from an eye tracking study. *Educ. Psychol. Rev.* 33, 91–107. doi: 10.1007/s10648-020-09535-z

TYPE Original Research
PUBLISHED 24 April 2023
DOI 10.3389/feduc.2023.1062053

OPEN ACCESS

EDITED BY
Maike Schindler,
University of Cologne, Germany

REVIEWED BY
Sarah Hansen,
Columbia University, United States
DaiWai Olson,
University of Texas Southwestern Medical Center, United States

*CORRESPONDENCE
Marc Rodemer
✉ marc.rodemer@uni-due.de

RECEIVED 05 October 2022
ACCEPTED 06 April 2023
PUBLISHED 24 April 2023

CITATION
Rodemer M, Karch J and Bernholt S (2023) Pupil dilation as cognitive load measure in instructional videos on complex chemical representations.
Front. Educ. 8:1062053.
doi: 10.3389/feduc.2023.1062053

Pupil dilation as cognitive load measure in instructional videos on complex chemical representations

Marc Rodemer[1]*, Jessica Karch[2] and Sascha Bernholt[3]

[1]Department of Chemistry Education, University of Duisburg-Essen, Essen, Germany, [2]Department of Chemistry, Tufts University, Medford, MA, United States, [3]Department of Chemistry Education, Leibniz-Institute for Science and Mathematics Education, Kiel, Germany

This secondary analysis of an earlier eye-tracking experiment investigated how triangulating changes in pupil dilation with student-self reports can be used as a measure of cognitive load during instructional videos with complex chemical representations. We incorporated three signaling conditions, dynamic, static and no signals, into instructional videos to purposefully alter cognitive load. Our results indicate that self-reported extraneous cognitive load decreased for dynamic signals compared to static or no signals, while intrinsic cognitive load was not affected by the signaling condition. Analysis of pupil dilation show significantly larger pupils for dynamic signals as compared to the other two conditions, suggesting that when extraneous cognitive load decreased, students still engaged cognitively with the task. Correlation analyses between measures were only significant for pupil dilation and extraneous cognitive load, but not pupil dilation and intrinsic cognitive load. We argue that beneficial design choices such as dynamic signals lead to more working memory capacity that can be leveraged toward learning. These findings extend previous research by demonstrating the utility of triangulating self-report and psychophysiological measures of cognitive load and effort.

KEYWORDS
eye-tracking, pupillometry, cognitive load, instructional videos, chemistry, representations

1. Introduction

To develop learning materials that align with insights from cognitive science, theories of human cognitive architecture are used to shape instructional approaches. Particularly in STEM, where the subject matter gets increasingly abstract and complex, effective learning materials are indispensable. One consensus framework guiding instructional design is Cognitive Load Theory (CLT) which proposes that learning occurs when information is initially processed in working memory and subsequently stored in long-term memory (Sweller et al., 1998, 2019). The mental effort expended in working memory is referred to as cognitive load, and during learning this load can be induced by the difficulty of the task (referred to as intrinsic cognitive load, or ICL) or by its design (referred to as extraneous cognitive load, or ECL; Sweller et al., 2019). Because working memory is limited in capacity and duration, learning is impeded when working memory capacity is exceeded, i.e., when one experiences excessive cognitive load. One goal of CLT-informed instructional design is to minimize ECL in order to keep enough working memory resources free for managing ICL of the material to be learned.

Following STEM researchers' and teachers' interest in supporting student learning by altering and optimizing instructional design, we investigated the impact of several design choices on students' cognitive load. We designed instructional videos on organic chemistry reaction mechanisms, because (1) small alterations can be made to videos to detect differences while keeping the overall instruction constant, and (2) reaction mechanisms are known to be visually and conceptually demanding and thus difficult to learn for students (for reviews, see Gilbert, 2005; Graulich, 2015; Daniel et al., 2018). One main student challenge in organic chemistry involves understanding the domain-specific representations and linking them to the underlying chemical concepts. Students often struggle to identify the relevant entities (Rodemer et al., 2020), which induces cognitive load (Rodemer et al., 2022). Since these chemical representations are intrinsically complex, unnecessary cognitive load might be counteracted with reducing extraneous load by optimized instructional design. To examine the impact of design on load, we chose different signaling techniques derived from multimedia learning principles (de Koning et al., 2009; Mayer, 2014; van Gog, 2014). By guiding students' attention to relevant parts of the learning material, signals facilitate comprehension and reduce cognitive load. Specifically, we compared how cognitive load is influenced by three signaling conditions: sequential signaling (dynamic), permanent signaling (static), and no signaling (control).

Cognitive load can be assessed by using psychophysiological and self-reported measures. A well-known and reliable indicator for cognitive load is pupil dilation, which can be measured with an eye tracker. Pupillometry has been used extensively to investigate cognitive load in different learning scenarios (for reviews, see Beatty and Lucerno-Wagoner, 2000; Just et al., 2003; van der Wel and van Steenbergen, 2018). However, little work has examined the influence of different types of load on pupil dilation. The present study is a secondary analysis from our prior eye-tracking experiment focusing on the impact of signals on learning outcomes, cognitive load and attention (Rodemer et al., 2022). In this report, we present the first analysis of pupil diameter and its relationship to the previously reported cognitive load self-reports.

2. Theoretical background

2.1. Cognitive load theory

Cognitive Load Theory describes that learning capability is influenced by human cognitive architecture. More specifically, learning capability is limited by the capacity of human working memory (Sweller et al., 1998, 2011, 2019; van Merriënboer and Sweller, 2005). The amount of information that can be processed simultaneously in working memory restricts the amount of information that can be learned, i.e., information that can be stored in long-term memory. The limitation of working memory accounts especially for novel information that is obtained through sensory systems, since this information must be ordered and integrated (van Merriënboer and Sweller, 2005). The acquisition of expertise, or in other words, learning, is hindered when working memory capacity is exceeded (Sweller et al., 2019).

Regarding the two types of cognitive load, ICL is determined by the expertise of the learner and their interaction with the given nature of the learning material (van Merriënboer and Sweller, 2005). It is caused by the amount of information that must be processed simultaneously in working memory, i.e., ICL depends on the extent of element interactivity of the learning material. The larger the number of interacting elements, the more difficult the given content is understood. In order to facilitate understanding, these interacting elements need to be incorporated into cognitive schemata, which are acquired over time through experience with subject material. Thus, ICL of a task or material decreases with expertise in a specific domain (van Merriënboer and Sweller, 2005). With a specific learning goal and learning task at hand, ICL cannot be altered purposefully by instructional interventions (van Merriënboer and Sweller, 2005). In contrast, ECL does not contribute to load necessary for understanding the material at hand (van Merriënboer and Sweller, 2005). ECL is induced by sub-optimal design choices, where a learner has to search for relevant information, or by triggering weak problem-solving methods (van Merriënboer and Sweller, 2005). Hence, ECL can be altered purposefully by instructional interventions.

Intrinsic cognitive load and ECL have an additive relationship to each other. If one load is exceeded, working memory capacity is exceeded in total, resulting in impeded learning (Paas et al., 2003; Cowan, 2010). If a task is perceived as easy, i.e., ICL is low, then a high ECL might be manageable for a learner, since the overall working memory capacity is kept within its limits. However, if ICL is high, ECL must be decreased in order for a learner to work through a task without cognitive overload (Kalyuga, 2011; Sweller et al., 2019). Hence, the goal of well-designed instructional material is to reduce ECL so that available cognitive resources can be fully devoted to the actual learning process (Mayer, 2005, 2021).

2.2. Multimedia design principles to reduce cognitive load

Based on CLT, the Cognitive Theory of Multimedia Learning proposes several design principles in order to manage cognitive load effectively (Mayer, 2005, 2021). Multimedia formats such as instructional videos utilize both the auditory and visual sensory channels, which has specific implications for designing these learning materials. Building upon this dual-channel assumption (Clark and Paivio, 1991), the CTML puts forward that auditory and visual information must first be integrated in working memory before they can be stored in long-term memory (Mayer, 2021). In line with CLT and CTML, attention that is available for each of these two separate information processing channels is limited (limited-capacity assumption; Mayer, 2014). Multimedia learning material is considered effective when each channel is addressed in its natural form, i.e., when images or representations are seen and when sounds are heard (Mayer, 2014). To leverage learning, the modality principle suggests verbal explanations better complement visual stimuli as opposed to displaying text on screen (Low and Sweller, 2014). Other well-researched principles for reducing ECL are summarized as follows (Mayer and Fiorella, 2014): The coherence principle declares that task-irrelevant details, such as additional texts or decorative pictures, should be excluded. The redundancy principle emphasizes that information that is simultaneously provided through multiple sensory channels places additional cognitive load on the learner, e.g., by providing a verbal narration and printed text. The spatial and temporal contiguity principle suggests that corresponding words and pictures should be presented near to each other or simultaneously rather than separately or successively.

Beyond these guidelines for reducing ECL, a great body of research is concerned with the signaling principle (for meta-analyses; see Richter et al., 2016; Xie et al., 2017; Schneider et al., 2018; Alpizar et al., 2020). The signaling principle states that a visual cue or highlight that emphasizes relevant parts of the learning material reduces ECL by guiding attention, particularly when the amount of information is difficult to change. Signals can appear as a circle, arrow, or by coloring specific parts. A visual signal is known to support a learner to focus on relevant features of a display. The underlying mechanism is that cognitive resources that might otherwise be directed toward visual search are freed up (de Koning et al., 2009).

2.3. Measuring cognitive load: self-reports and pupillometry

Several approaches to measuring cognitive load have been proposed. These approaches are based either on subjective judgments or on objective measurements, and thus address load either directly, e.g., by asking learners to rate their perceived mental load, or indirectly, e.g., by using indicators that are thought to reflect learners' mental load, such as performance (Klepsch et al., 2017). Generally, the different approaches all show strengths and weaknesses (see Brünken et al., 2010). In educational research, subjective ratings of cognitive load are the most frequently used approach (e.g., Schmeck et al., 2015; Krieglstein et al., 2022). In these approaches, the learner is asked, in most cases retrospectively, to rate the perceived amount of cognitive load on a Likert scale while working on a task. Generally, this approach is considered beneficial due to its economy and flexibility. In addition, the retrospective rating does not disturb the learning process and impose load by itself, which may be the case in other approaches such as dual-task measures (Brünken et al., 2010). A recent meta-analysis concluded that self-reports of perceived cognitive load also are a valid and reliable measure (Krieglstein et al., 2022). However, there is empirical evidence that the rating of cognitive load depends on certain personal and situational aspects, such as the timing of the measurement (Brünken et al., 2010) and subjective internal standards for evaluating current load state (Klepsch et al., 2017). Furthermore, multidimensional measures of cognitive load often show significant correlations between different types of cognitive load (e.g., ICL and ECL), which seems inconsistent with the additivity hypothesis of the cognitive load theory (Krieglstein et al., 2022).

Another stream of research is concerned with small changes in pupil diameter that are attributed to reflect changes in brain activity, or, more specifically, human cognition (Beatty and Lucerno-Wagoner, 2000; Just et al., 2003; van der Wel and van Steenbergen, 2018). In this stream, pupil dilation has been used as a proxy measure for many cognitive processes, including arousal, attention, and cognitive load (Stanners et al., 1979; Klingner et al., 2010; Kang et al., 2014; Miller and Unsworth, 2020). Although this relationship between pupil dilation and cognitive effort was first reported over 100 years ago (e.g., Löwenstein, 1920), it was popularized as a systematic course of study with seminal studies in the mid-1960s, which demonstrated that an increase in pupil size compared to baseline, up to 0.5 mm, could be discretely correlated with mental effort exerted in increasingly complex numerical recall tasks (Hess and Polt, 1964; Kahneman and Beatty, 1966). Recent neurobiological studies suggest that this effect is due to activation of the noradrenergic system's locus coeruleus, which is activated by stress, and may also play a role in memory consolidation (Beatty and Lucerno-Wagoner, 2000; Laeng et al., 2012; van der Wel and van Steenbergen, 2018).

Studies on task-evoked pupillary responses (TEPR) focus on how changes in attention or cognitive effort during a task can be measured through changes to pupil size compared to baseline. This response can be isolated through careful control of the environment, e.g., controlling external stimuli such as change in brightness or excessive movement that can induce a change in pupil size, and careful design of the experiment to reduce the number of conflicting cognitive signals (Beatty and Lucerno-Wagoner, 2000; Karch, 2018).

The relationship between experimental design and the nature of what is being assessed through pupil dilation is not straightforward. Many studies correlate pupil dilation with task demand, e.g., cognitive load, particularly for simple tasks such as arithmetic, repeating back an increasingly long stream of numbers or letters, or entering a difficult password (Hess and Polt, 1964; Kahneman and Beatty, 1966; Klingner, 2010; Krejtz et al., 2018; Abdrabou et al., 2021). However, a recent meta-review of TEPR studies suggests that this relationship is more complicated, and that pupil dilation can better be understood as cognitive *effort* rather than task demand (van der Wel and van Steenbergen, 2018). Thus, it is crucial to understand how a participant may be experiencing a task in order to interpret their pupil dilation, because more novice performers in a task may have higher pupil dilations to reflect that they need to put in more effort to grapple with the task, and more expert performers may have smaller dilations due to the fact that they need to exert less effort (Ahern and Beatty, 1979; Szulewski et al., 2017; van der Wel and van Steenbergen, 2018; Zhou et al., 2022).

There have been several promising studies that it is possible to ascribe meaning to pupil dilations collected *in situ*, e.g., while one is engaged in a task, to understand how engaging with the task involves cognitive effort (e.g., Palinko et al., 2010; Krejtz et al., 2020; da Silva Castanheira et al., 2021; Shechter and Share, 2021). However, few have tried to make claims about the nature of the cognitive load that induces this effort, in part because of the difficulty associated with interpreting psychophysiological signals (Cacioppo and Tassinary, 1990). Some have done so through deliberate experimental design. For example, Foroughi et al. (2017) found that pupil size decreased as participants completed multiple trials of an experiment, suggesting they automatized the process. Shechter and Share (2021) conducted word recognition experiments, finding significantly larger relative changes in pupil size for stimuli associated with higher cognitive effort. Another way to investigate pupil dilations may be to triangulate other sources of data, such as gaze data (e.g., Klingner, 2010; Karch et al., 2019; Miller and Unsworth, 2020), spatio-temporal sensory cues (e.g., Sharma et al., 2021), interviews (e.g., Pomerleau-Turcotte et al., 2021), motivational manipulation by task-switching (da Silva Castanheira et al., 2021), microsaccadic responses (Krejtz et al., 2020), and through probes mid-task (Franklin et al., 2013) to try to understand the underlying cognitive process reflected in the pupil dilation.

2.4. The present study

The goal of the present experiment was twofold. The first was a conceptual goal. We wanted to understand how different design choices for signaling during instructional videos impacted the cognitive load students experienced while watching these videos (RQ1). While there is a large body of research supporting the cognitive benefit of signals, most of this evidence is based on learning outcomes. Additionally, many studies make use of rather simple tasks that require rapid mental operations in working memory (e.g., arithmetic or memory scanning) and/or that can be solved without substantial prior knowledge but on basis of the given instruction. The first research question of this study is:

RQ1: Under which condition is either students' ICL or ECL reduced while watching instructional videos with either dynamic, static, or no signals?

Based on the literature, we expect to reduce ECL by providing signals in a descending order from control to static to dynamic signaling, e.g., that tasks with the control signal will result in the highest ECL, whereas tasks with dynamic signaling will have the lowest ECL. Furthermore, we hypothesize that based on our task design, ICL will be kept constant across signaling conditions (Richter et al., 2016; Xie et al., 2017; Schneider et al., 2018; Alpizar et al., 2020).

The second was a methodological goal. Although pupillometry can potentially offer an *in situ* method to examine how cognitive load changes over time, few studies have looked at change in pupil dilation while watching instructional videos (Huh et al., 2019), in part because pupil signals can be challenging to isolate and interpret. Additionally, the relationship between pupil dilation and different types of cognitive load is unclear. Traditional TEPR studies focus on the relationship between pupil dilation and task difficulty, e.g., ICL (e.g., Hess and Polt, 1964; Kahneman and Beatty, 1966; Szulewski et al., 2017). Some studies have started to look at how altering the design of a task to provide visual supports impacts cognitive load, e.g., focusing on the relationship between pupil dilation and ECL (Zheng and Cook, 2012; Kruger et al., 2013). However, neither of these studies conducted a targeted study on the relationship between ECL and pupil size, but rather looked at the effect on cognitive load as a whole. Mitra et al. (2017) used CLT to show that it is possible to use pupillary responses to infer the extent to which students experience different types of cognitive load, but their study was conducted using fairly straightforward tasks such as question comprehension or mental math. Thus, we wanted to understand how pupil dilations change when the ECL of an authentic instructional task is altered, due to modifying the signaling condition but not the difficulty of the tasks (RQ2). By triangulating self-report measures and psychophysiological measures of cognitive load, our goal is to contribute to making pupillometry a more useful and interpretable measure for educational research. Thus, the second research question is:

RQ2: Does pupillometry indicate differences in pupil diameter when altering extraneous load across experimental conditions?

Our hypothesis is that pupil diameter is affected by different extraneous load conditions. We predict that as extraneous load goes down across the three signaling conditions, we will see a corresponding decrease in pupil dilation.

3. Materials and methods

3.1. Sample and study design

The study presented here is a re-analysis of prior work from the first and last author, which focused primarily on how the signaling conditions in the instructional videos impacted students' attention, self-reported cognitive load, and learning outcomes (Rodemer et al., 2022). In this study, 28 undergraduate chemistry students (50% female, 50% male; 0% nonbinary) from a German university participated on a voluntary base in winter semester 2019. Participants were currently enrolled in an introductory general chemistry course to ensure that they had sufficient prior knowledge to potentially understand the rather complex chemical reactions that were presented in our instructional videos. All participants had normal or corrected-to-normal vision. None of the participants reported on either color vision deficiency or specific learning disabilities (e.g., dyslexia) that might have impacted their processing of the videos or their cognitive load.

A 1×3 within-subject design was employed in which the instructional videos were manipulated according to three different *signaling conditions* (i.e., *no signaling* vs. *static signaling* vs. *dynamic signaling*). Each participant watched three videos in a constant video order but received each video including one of the three signaling conditions. To control for potential sequencing effects, each of the three signaling conditions were presented according to a counterbalanced 3×3 Latin Square design to evaluate potential effects of treatment position and video content on the dependent variables (Tabachnick and Fidell, 2007; see also Figure 1). Participants were randomly assigned to one of the three treatment sequences, which were implemented as a between-subject factor.

3.2. Material and measures

3.2.1. Instructional videos

Three instructional videos covering introductory organic reaction mechanisms at the university level were developed (Eckhard et al., 2022). Each video focuses on one of three chemical factors that influence reaction speed, namely leaving group ability (video 1), substrate effects (video 2), and nucleophilic strength (video 3). Overall, the difficulty of each video was comparable since the chemical factors chosen for each example can be understood independently from each other and do not built upon each other. To keep the design of the videos constant, representations on the display were arranged the same way and verbal explanations that accompanied the task followed the same structure. Videos had a length of approximately 5 min each (for German (original) and English (translated) videos, see: https://osf.io/r4sx3/).

Each video was presented as a case comparisons of nucleophilic substitution reactions. This task format is common in chemistry entailing complex representations, such as structural formulas and electron-pushing arrows (Caspari et al., 2018; Graulich and Schween, 2018; Bodé et al., 2019). Students needed to compare commonalities and differences between the representations, connect these features to chemical factors from the verbal explanation, and critically weigh the factors in terms of their influence on the reaction speed. The corresponding verbal explanations were narrated in line with

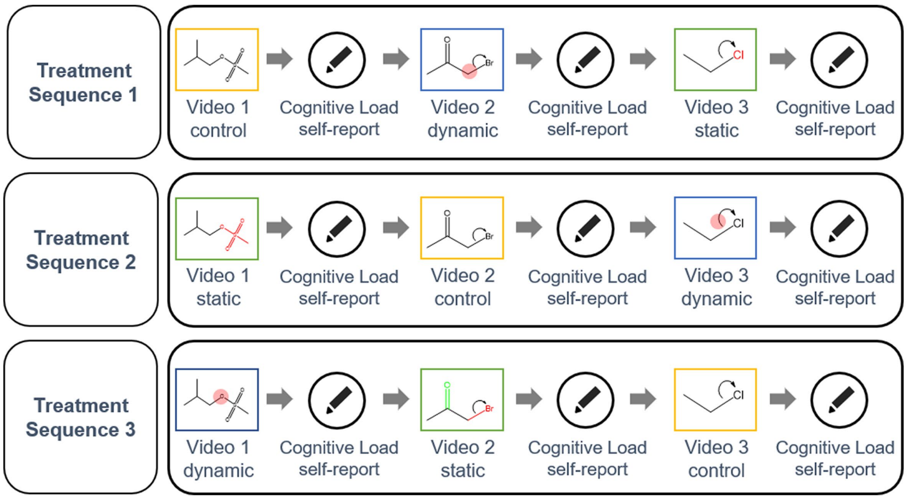


FIGURE 1
Experimental 3x3 Latin Square design showing signaling condition, instructional video and treatment sequence. The boxes display a screenshot from the chemical representations that were explained in the videos and the design of the signaling conditions. All participants received the instructional videos in the same order with regards to content, while the order of the signaling condition differed according to their treatment sequence.

recommendations based on the modality principle (Ayres and Sweller, 2014). The explanations followed a step-by-step structure that are commonly used in worked examples (Renkl, 2014). The only aspect in which the videos differed from each other were the example reactions that were chosen to highlight different factors that influence reaction speed. The structure of the explanation was kept comparable across instructional videos.

Concerning the experimental factor *signaling condition*, either no signals (i.e., control condition), static signals (i.e., permanent coloring of specific representational features), or dynamic signals (i.e., a sequential red dot) were added to the videos. The dynamic signal was embedded when the narration mentioned specific relevant features of the representations, lasting anywhere from single words to several consecutive sentences. For each instructional video (i.e., Videos 1, 2, and 3), the narrated explanation was identical in all signaling conditions but differed between instructional videos based on the content they present.

3.2.2. Pupil diameter recording and data pre-processing

The instructional videos were presented on a 24-inch screen with a 1920 × 1080 pixel resolution using the software Tobii Pro Lab. Participants sat in front of the screen at approximately 60-cm distance with headsets to follow the verbal explanation of the videos without distractions. Participants' pupil diameters were recorded using a Tobii Pro Spectrum, an eye tracking device with a 200 Hz sampling rate which estimates true pupil size based on the participants' distance from the eye tracker and shape of their cornea (Karch, 2018). The system was calibrated using 9-point calibration and subsequent validation. The calibration accuracy was below 0.5° for all participants ($M = 0.30°$, $SD = 0.21°$).

To prepare pupillometric data for analysis, raw data with Tobii I-VT Fixation Filter with a threshold of 30°/s were exported from Tobii Pro Lab. Raw data were uploaded to and processed in RStudio. Following Mathôt's (2018) guidelines on pre-processing pupil data and adapting code from the second author (Karch, 2018), blinks were removed by calculating a velocity profile to identify when there were rapid changes in pupil size, indicating that the eyes closed, and removing points that fell outside of the threshold of three standard deviations from the median velocity (Leys et al., 2013; Kret and Sjak-Shie, 2019). Then data were smoothed using a rolling average over three data points (a window of 15 ms) to remove potential noise at very high frequencies from instrument error. Finally, for each video, baseline values calculated based on the median of the first ten samples were subtracted from all pupil size values to give dilation data. These baseline-subtracted dilation values were then used for all statistical analyses described below (processing code can be accessed online at: https://osf.io/r4sx3/).

3.2.3. Cognitive load measures

We used the established self-report scales by Klepsch et al. (2017) to measure intrinsic and extraneous cognitive load. Participants rated their perceived cognitive load on a 7-point rating scale (1 = low, 7 = high) immediately after each video. The cognitive load items were presented on the computer screen and were read aloud by the test supervisor. To adapt the measure according to the context of the study, the wording in the items were changed from "task" to "video." Cronbach's α indicated a sufficiently high reliability of the two scales ($\alpha_{ICL} = 0.91$; $\alpha_{ECL} = 0.85$).

3.2.4. Procedure

The study followed ethical standards recommended by the German Research Foundation: Upon arrival, participants were fully informed about the voluntary nature, goals, process, and data handling of this study. All participants signed a written informed consent and were aware that they could withdraw their consent at any time.

The study was performed in single sessions of 1.5 h in a light-controlled environment. After completing a pen and paper questionnaire about demographics, participants were familiarized with the eye tracker and the calibration procedure. Calibration was repeated until high accuracy was reached. Then, participants were instructed to watch the three instructional videos carefully and that they could not pause or rewind. In between each video, the cognitive load items were asked. Once the instruction was completed, participants received a monetary compensation. The procedure was kept constant for all participants.

3.2.5. Data analyses

Our analyses of variance were focused on the effects of the experimental signaling conditions (within-subject measure), potential differences across the three instructional videos, and also included the sequence of experimental conditions (see Figure 1) as a between-subject factor. To answer our research questions, we were most interested in the main effects of the three signaling conditions. Position and sequence effects were investigated to control for potential content or carryover effects across conditions in our within-subject design. Post-hoc pairwise comparisons were performed using Benjamini-Hochberg adjustment. As measure for the effect size (partial) η^2 and the correlation coefficient r are reported, where values are interpreted according to Cohen (1988). Statistical analyses were performed using R Version 4.0.4 and several packages, notably 'tidyverse', 'ExpDes', 'rmcorr', and 'lme4' (Bates et al., 2015; Bakdash and Marusich, 2017; Wickham et al., 2019; Batista Ferreira et al., 2021; R Core Team, 2021).

4. Results

4.1. Self-reports of cognitive load

As expected, our analysis of ICL showed no main effect for the experimental factor *signaling condition* ($F(2,83) = 0.26$, $p = 0.769$, $\eta^2 = 0.05$; Figure 2, left). Further, we found no significant effect for the *instructional video* ($F(2,83) = 2.24$, $p = 0.113$, $\eta^2 < 0.01$), whereas a significant effect was present for the factor *treatment sequence* ($F(2,83) = 5.69$, $p = 0.005$, $\eta^2 = 0.13$). Regarding the treatment sequence, pairwise comparisons indicated that Sequence 2 (dynamic/static/control) showed significantly lower ICL compared to both Sequence 1 (control/dynamic/static, $p = 0.003$) and Sequence 3 (static/control/dynamic, $p = 0.048$). Sequences 1 and 3 did not differ significantly ($p = 0.319$).

As predicted, the analysis of ECL showed a significant main effect for the experimental factor *signaling condition* ($F(2,83) = 8.89$, $p < 0.001$, $\eta^2 = 0.19$; Figure 2, center). Pairwise comparisons indicated a significant lower ECL for the condition with dynamic signals compared to the condition with static signals or no signals in the control condition (both $p < 0.001$). There was no difference between the static and control condition ($p = 1.00$). Furthermore, we found no significant effects for the factor *instructional video* ($F(2,83) = 1.54$, $p = 0.221$, $\eta^2 = 0.04$) or the factor *treatment sequence* ($F(2,83) = 2.88$, $p = 0.062$, $\eta^2 = 0.07$).

4.2. Pupil diameter

With regard to mean pupil dilation values, the analysis showed a significant main effect for the experimental factor *signaling condition* ($F(2,83) = 3.24$, $p = 0.045$, $\eta^2 = 0.08$; Figure 2, right), no significant effect for the *instructional video* ($F(2,83) = 1.72$, $p = 0.186$, $\eta^2 = 0.04$), and no significant effect for the factor *treatment sequence* ($F(2,83) = 1.04$, $p = 0.355$, $\eta^2 = 0.03$). Although the average measures for all pupil data suggest that during task participants' eyes were constricted compared to baseline, pairwise comparisons indicated a significant larger relative dilation for the condition with dynamic signals compared to the condition with no signals in the control condition ($p = 0.001$). There was no significant difference between the static and control condition ($p = 0.065$) and between the static and the dynamic condition ($p = 0.195$).

To gain more fine-grained insights into the processes of the video consumption, students' pupil dilation has been analyzed across time for each of the videos. Figure 3 illustrates the time course of the average pupil dilation across participants for each of the three videos and separated by treatment condition. Peaks and valleys represent changes in pupil dilation over time, where peaks represent instances of higher cognitive load. These graphs show that the dilatory response to videos 1 and 3 were consistently higher in the dynamic condition compared to the control and static condition, and that the control condition was the lowest, while the mean dilations across time for video 2 tended to be more similar. Additionally, the shapes of the graphs, i.e., where there tended to be peaks and valleys, were relatively similar across all three conditions, suggesting that students may have experienced stimuli that induced cognitive load at similar points. This is what we would anticipate, as the scripts and video were identical across all three conditions. These time course graphs provide

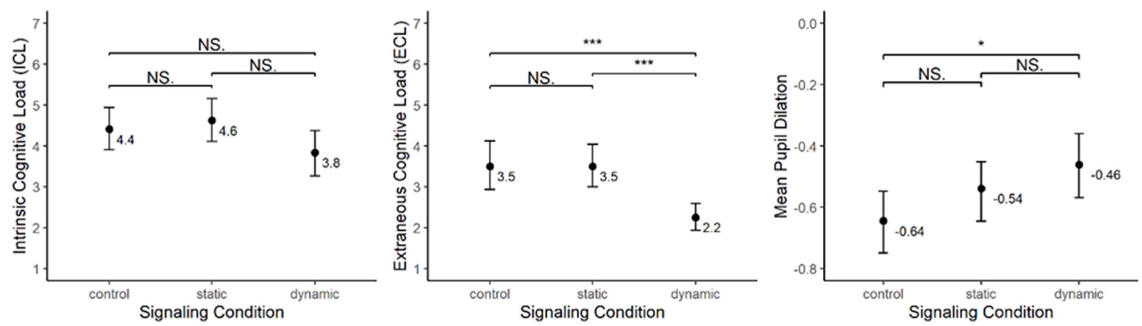


FIGURE 2
Means for intrinsic cognitive load (left), extraneous cognitive load (center), and pupil dilation (right) by signaling condition. Points indicate mean-values, error bars 95% confidence intervals. Results of pairwise comparisons are indicated by significance levels (NS. p>0.05; *p<0.05; **p<0.01; ***p<0.001).

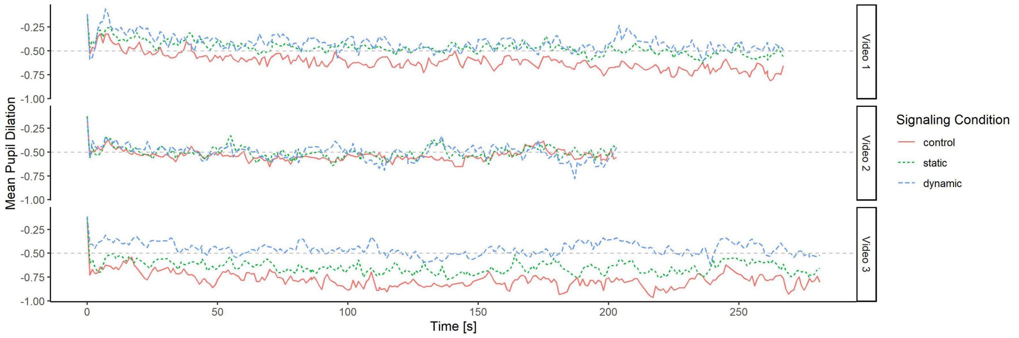


FIGURE 3
Pupil dilation over time per video and across different signaling conditions.

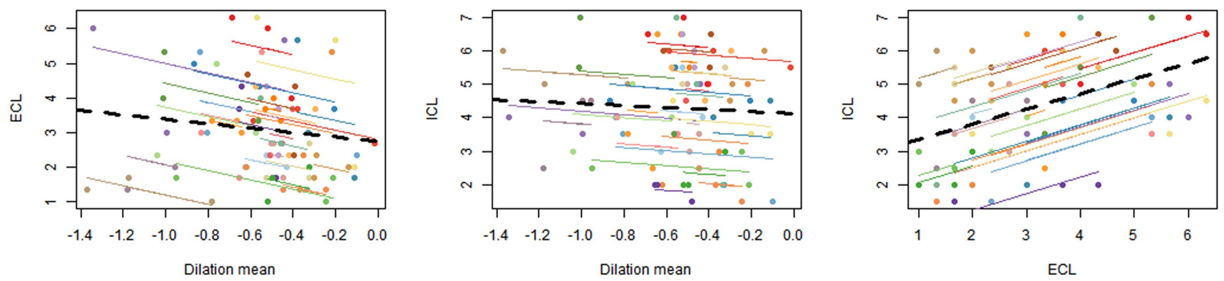


FIGURE 4
Repeated measures correlation plots for the association between pupil dilation and extraneous cognitive load (ECL; Left) and intrinsic cognitive load (ICL; Center), respectively, as well as between ECL and ICL (Right). Each participant provides three data points (one per video) that are shown in a different color. The colored lines show repeated measurement correlation fits for each participant. The black line indicates the overall regression line.

additional qualitative evidence that the mean pupil dilations shown in Figure 2 (right) reflected differences that were maintained across the entire course of each instructional video.

4.3. Correlation between cognitive load and pupil dilation

Repeated measures correlation coefficients for pairwise correlations between self-reported cognitive load scales (ICL and ECL) and pupil dilation were calculated to analyze the relationship between these measures. Findings indicate a negative association between ECL and pupil dilation, r_{rm} (55) = −0.25, 95% CI [−0.48, −0.02], $p = 0.06$ (Figure 4, left), i.e., when students report higher extraneous cognitive load after watching the video, their mean pupil dilation is more negative, indicating smaller pupil size. The association between ICL and pupil dilation is also negative, but smaller and not significant (r_{rm} (55) = −0.10, 95% CI [−0.30, 0.17], $p = 0.46$; Figure 4, center). The association between ICL and ECL is positive (r_{rm} (55) = 0.53, 95% CI [0.33, 0.72], $p < 0.01$; Figure 4, right).

5. Discussion

This secondary analysis followed a conceptual and a methodological goal. The first goal was to investigate how different types of signaling impacted students' cognitive load while watching instructional videos containing complex chemical representations. The second goal was to examine the relationship between pupil dilation and self-reports while altering different types of cognitive load. To approach these goals we implemented dynamic, static, or no signaling in instructional videos and recorded pupil dilations with an eye-tracker as well as collected self-reports on intrinsic and extraneous cognitive load with an established questionnaire.

The analysis of ICL self-reports showed no main effect for the experimental factor signaling condition. This result was expected since the difficulty of each instructional video was kept comparable. A main effect was found for the treatment sequence, indicating that participants perceived the instructional videos to be easier when they received them in the order dynamic–static–control. This finding may possibly be attributed to fading-out support over time—an instructional principle that is well-known in research concerning worked examples (Renkl, 2014). In such a fading procedure, full support is provided in the first example. Then, in the following examples, the amount of support decreases until only the problem that is to be solved is left.

Results of ECL self-reports showed a significant reduction for dynamic signals as compared to static or no signals. Consistent with CLT and CTML (Sweller et al., 2019; Mayer, 2021), the reduction of ECL through dynamic signals can be attributed to a reduction of search space. Showing a dynamic signal facilitated information selection from the visual representations. Furthermore, the dynamic signal supported integrating the audible explanation and the visual representation. Based on the dual-coding assumption and the CTML, we argue that the dynamic signal from our instructional video supports the integration of the auditory and visual information in working memory by increasing attention to relevant entities, and, thus freeing up working memory capacity that otherwise would be attributed to searching the relevant representations that are mentioned in the explanation. Considering the intrinsic complexity of the chemical representations, reducing unnecessary load might support students in overcoming their difficulty in connecting these representations with the underlying concepts (Graulich, 2015).

Given this finding, we would have expected there to be a corresponding decrease in cognitive load as measured by pupil dilation. However, results showed significantly larger pupils in the dynamic signaling condition as compared to the control condition without signals, but not significantly different for the comparisons dynamic—static and static—control. This result is surprising because we anticipated that a dynamic signal would decrease cognitive load, and thus lead to smaller pupil dilation (Hess and Polt, 1964; Klingner et al., 2010). A possible explanation might be that dynamic signaling increased cognitive processing, e.g., working memory allocated to productive mental effort, as opposed to cognitive load, e.g., working memory allocated to deal with a task (Krejtz et al., 2020; Shechter and Share, 2021). Comparing both results, the reduction of ECL by self-reports and the increase of pupil dilation, supports the interpretation of increased cognitive processing. When ECL is reduced and ICL stays constant, more working memory capacity can be directed toward (productive) mental effort, such as cognitive schemata formation, which is in line with findings described in TERP-literature (Mitra et al., 2017; van der Wel and van Steenbergen, 2018; da Silva Castanheira et al., 2021; Shechter and Share, 2021; Zhou et al., 2022). Another explanation might be that dynamic signals increased curiosity because of their movement, which would also be reflected in pupil diameter changes (van der Wel and van Steenbergen, 2018). Although we applied a tight research design varying only one factor, we cannot rule out this explanation because we did not collect additional affective variables. Consequently, further research is needed to inform a valid interpretation of findings based on changes in pupil diameter in the context of multimedia learning and complex, domain-specific representations.

One limitation of studies with pupillary data is that they cannot be interpreted in isolation, because pupillary signals may have many confounding sources. However, secondary sources of evidence can be used to support interpretation of pupillary data (e.g., Franklin et al., 2013; Krejtz et al., 2020; Miller and Unsworth, 2020; da Silva Castanheira et al., 2021). Our first secondary source of evidence are the self-reports of cognitive load, as discussed above. Our second source is evidence from two earlier studies from our research group that found gains in overall learning performance and retention moderated through dynamic signals in instructional videos (Rodemer et al., 2021, 2022). In the case of the control condition without signaling, ECL was higher and pupil dilation was lower, indicating that cognitive resources might be occupied by a visuospatial searching process. Without appropriate support, cognitive resources may have been overloaded, leading to participants' disengagement with the instruction, which is reflected in smaller pupil size and thus less mental effort (Peavler, 1974; Krejtz et al., 2018; van der Wel and van Steenbergen, 2018). In the case of the dynamic signaling, ECL was lower and pupil dilation was higher, and learning gains were increased, suggesting that the additional cognitive effort indicated by pupil dilation was a result of productive mental effort that led to these increased learning gains (Mitra et al., 2017; van der Wel and van Steenbergen, 2018).

Repeated measures correlation analyses show a significant correlation between pupil dilation and ECL but not ICL. This suggests the perceived inherent difficulty of the videos to be unrelated to the extent of cognitive processing, while the video design, e.g., the signaling condition, seems to be the more important factor, at least in the present study. Mitra et al. (2017) showed that pupils dilated to different types of cognitive load. In their study, they altered the intrinsic difficulty (ICL) of the tasks while keeping the extraneous difficulty (ECL) constant. However, their study used very simple tasks that are hardly comparable to the rather complex instructional videos we used in our experiment, since the chemical representations presented require specific domain-specific understanding which is not the case for the graphs used in the study by Mitra and colleagues. Although our results indicate a crucial role of the instructional design that takes the extraneous difficulty into account, more systematic research is needed to further investigate the relations between different types of cognitive load and pupil dilation, particularly during domain-specific learning tasks.

6. Implications for practice and research

When designing this experiment, we argued that instructional design should be modified with the goal to reduce extraneous cognitive load, e.g., by implementing dynamic signals. The results from this study suggests that not only do dynamic signals reduce ECL, this reduction may free up enough mental resources that students have a larger capacity to grapple with the task itself or with learning processes. This is suggested by the presence of larger pupil sizes during tasks with lower reported ECL, suggesting that students were still cognitively engaged and putting mental effort into the task. This has several implications for practice. First, implementing dynamic signaling in instructional videos may support student learning in the class. Second, although the videos in our study were designed to all have the same relative level of difficulty, it is possible that the resources freed up by reducing ECL may free up space for higher levels of ICL. That is, dynamic signaling may be a useful scaffold when instructors introduce more intrinsically difficult tasks. Third, our study provides support for a transfer of the *fading principle* to the application of signaling in instructional videos. The fading principle describes gradually fading support over time which was originally described in Renkl's (2014) theory of example-based learning. Finally, although our study focuses on the use of dynamic signaling in organic chemistry instructional videos, the theoretical foundation of the work is not drawn from chemistry but rather CMTL, thus it may be possible that our findings on the effect of dynamic signaling in instructional videos may be applicable to other domains.

With regards to research, we demonstrated the utility of triangulating findings from self-report cognitive load measures and pupillometric data. In particular, we showed that combining these two streams of data facilitated a more nuanced analysis of the possible effect of reducing ECL, e.g., freeing up resources for students to engage and provide effort in other ways when working with the task. Our research demonstrates promising potential that in combination with secondary data sources (self-reports and student outcomes), pupillary data can be meaningfully interpreted in more naturalistic and complex educational tasks, such as the case comparisons reported here. Future research should investigate pupil dilation by systematically varying stimuli with different levels of difficulty to induce different amounts of intrinsic cognitive load.

7. Conclusion

This study found that dynamic signals as compared to static or no signals reduced students' self-reported extraneous cognitive load without impacting intrinsic cognitive load during the consumption of instructional videos containing complex chemical representations. Furthermore, significant correlations were only found between pupil dilation and self-reported extraneous cognitive load, but not intrinsic cognitive load. Our results call for a stronger emphasis on instructional design to manage cognitive load. Based on the assumption that pupil dilation indicates mental effort, more systematic research is needed that investigates different types of cognitive load across tasks and instructions that vary in context and complexity.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MR: conceptualization, methodology, formal analysis, investigation, resources, data curation, writing—original draft, and visualization. JK: conceptualization, methodology, formal analysis, writing—original draft. SB: conceptualization, methodology, formal analysis, writing—review and editing, supervision, project administration, and funding acquisition. MR and JK contributed equally to the writing of the original draft. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft) under Grant 329801962.

Acknowledgments

We thank all the students who participated in our study. We also thank Julia Eckhard, Nicole Graulich, Gyde Asmussen, and Svea Hinrichsen.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdrabou, Y., Abdelrahman, Y., Khamis, M., and Alt, F. (2021). Think harder! Investigating the effect of Passwort strength on cognitive load during password creation. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery, 1–7.

Ahern, S., and Beatty, J. (1979, 4412). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science (New York, N.Y.)* 205, 1289–1292. doi: 10.1126/science.472746

Alpizar, D., Adesope, O. O., and Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educ. Technol. Res. Dev.* 68, 2095–2119. doi: 10.1007/s11423-020-09748-7

Ayres, P., and Sweller, J. (2014). "The Split attention principle in Mutlimedia learning," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer (Cambridge: Cambridge University Press), 206–226.

Bakdash, J. Z., and Marusich, L. R. (2017). Repeated measures correlation. *Front. Psychol.* 8:456. doi: 10.3389/fpsyg.2017.00456

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Batista Ferreira, E., Piscitelli Cavalcanti, P., and Alves Nogueira, D.. (2021). ExpDes: Experimental designs package [R package]. Version 1.2.2. Available at: https://CRAN.R-project.org/package=ExpDes (Accessed September 09, 2022).

Beatty, J., and Lucerno-Wagoner, B. (2000). "The pupillary system," in *Handbook of Psychophysiology*. eds. J. T. Cacioppo, L. G. Tassinary and G. G. Berntson. *2nd* ed (Cambridge: Cambridge University Press), 142–162.

Bodé, N. E., Deng, J. M., and Flynn, A. B. (2019). Getting past the rules and to the WHY: causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.* 96, 1068–1082. doi: 10.1021/acs.jchemed.8b00719

Brünken, R., Seufert, T., and Paas, F. G. (2010). "Measuring cognitive load," in *Cognitive Load Theory*. eds. J. L. Plass, R. Moreno and R. Brunken (Cambridge: Cambridge University Press), 181–202.

Cacioppo, J. T., and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *Am. Psychol.* 45, 16–28. doi: 10.1037//0003-066x.45.1.16

Caspari, I., Kranz, D., and Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.* 19, 1117–1141. doi: 10.1039/C8RP00131F

Clark, J. M., and Paivio, A. (1991). Dual coding theory and education. *Educ. Psychol. Rev.* 3, 149–210. doi: 10.1007/BF01320076

Cohen, J. (1988). *Statistical Power Analysis For the Behavioral Sciences*. *2nd* Edn. Hillsdale, NJ: L. Erlbaum Associates.

Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19, 51–57. doi: 10.1177/0963721409359277 da Silva Castanheira, K., LoParco, S., and Otto, A. R. (2021). Task-evoked pupillary responses track effort exertion: evidence from task-switching. *Cogn. Affect. Behav. Neurosci.* 21, 592–606. doi: 10.3758/s13415-020-00843-z

Daniel, K. L., Bucklin, C. J., Leone, E. A., and Idema, J. (2018). "Towards a definition of representational competence," in *Towards a Framework for Representational Competence in Science Education*. ed. K. L. Daniel (Cham: Springer International Publishing), 3–11.

Koning, B. B.De, Tabbers, H. K., Rikers, R. M. J. P., and Paas, F. (2009). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educ. Psychol. Rev.*, 21, 113–140. doi: 10.1007/s10648-009-9098-7

Eckhard, J., Rodemer, M., Bernholt, S., and Graulich, N. (2022). What do University students truly learn when watching tutorial videos in organic chemistry? An exploratory study focusing on mechanistic reasoning. *J. Chem. Educ.* 99, 2231–2244. doi: 10.1021/acs.jchemed.2c00076

Foroughi, C. K., Sibley, C., and Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology* 54, 1436–1443. doi: 10.1111/psyp.12896

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., and Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *Q. J. Exp. Psychol.* 66, 2289–2294. doi: 10.1080/17470218.2013.858170

Gilbert, J. K. (2005). "Visualization: a metacognitive skill in science and science education," in *Models and Modeling in Science Education: Vol. 1. Visualization in Science Education*. ed. J. K. Gilbert (Dordrecht: Springer), 9–27.

Graulich, N. (2015). The tip of the iceberg in organic chemistry classes: how do students deal with the invisible? *Chem. Educ. Res. Pract.* 16, 9–21. doi: 10.1039/C4RP00165F

Graulich, N., and Schween, M. (2018). Concept-oriented task design: making purposeful case comparisons in organic chemistry. *J. Chem. Educ.* 95, 376–383. doi: 10.1021/acs.jchemed.7b00672

Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science (New York, N.Y.)* 143, 1190–1192. doi: 10.1126/science.143.3611.1190

Huh, D., Kim, J.-H., and Jo, I.-H. (2019). A novel method to monitoring changes in cognitive load in video-based learning. *J. Comput. Assist. Learn.* 35, 721–730. doi: 10.1111/jcal.12378

Just, M. A., Carpenter, P. A., and Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theor. Issues Ergon. Sci.* 4, 56–88. doi: 10.1080/14639220210159735

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science (New York, N.Y.)* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Kalyuga, S. (2011). Cognitive load theory: how many types of load does it really need? *Educ. Psychol. Rev.* 23, 1–19. doi: 10.1007/s10648-010-9150-7

Kang, O., Huffer, K. E., and Wheatley, T. (2014). Pupil dilation dynamics track attention to high-level information. *PLoS One* 9, 1–6. doi: 10.1371/journal.pone.0102463.g001

Karch, J. M. (2018). "Beyond gaze data: Pupillometry as an additional data source in eye tracking," in *Eye tracking for the Chemistry Education Researcher*. eds. J. R. VandenPlas, S. J. R. Hansen and S. Cullipher. *1st* ed (Washington, DC: American Chemical Society), 145–163.

Karch, J. M., García Valles, J. C., and Sevian, H. (2019). Looking into the black box: using gaze and Pupillometric data to probe how cognitive load changes with mental tasks. *J. Chem. Educ.* 96, 830–840. doi: 10.1021/acs.jchemed.9b00014

Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.3389/fpsyg.2017.01997

Klingner, J. (2010). Measuring cognitive load during visual tasks by combining pupillometry and eye tracking. *Dissertation*. Stanford University, Stanford.

Klingner, J., Tversky, B., and Hanrahan, P. (2010). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology* 48, 323–332. doi: 10.1111/j.1469-8986.2010.01069.x

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., and Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One* 13:e0203629. doi: 10.1371/journal.pone.0203629

Krejtz, K., Duchowski, A. T., Zurawska, J., and Wichary, S. (2020). Pupillary and microsaccadic responses to cognitive effort and emotional arousal during complex decision making. *J. Eye Mov. Res.* 13:2. doi: 10.16910/jemr.13.5.2

Kret, M. E., and Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behav. Res. Methods* 51, 1336–1342. doi: 10.3758/s13428-018-1075-y

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., and Schneider, S. (2022). A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educ. Psychol. Rev.* 34, 2485–2541. doi: 10.1007/s10648-022-09683-4

Kruger, J.-L., Hefer, E., and Matthew, G. (2013). Measuring the impact of subtitles on cognitive load. In *ETSA '13 (Ed.), Proceedings of the 2013 Conference on Eye Tracking South Africa* (pp. 62–66). Association for Computing Machinery.

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Pers. Psychol. Sci. J. Assoc. Psychol. Sci.* 7, 18–27. doi: 10.1177/1745691611427305

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

Low, R., and Sweller, J. (2014). "The modality principle in multimedia learning," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer (Cambridge: Cambridge University Press), 227–246.

Löwenstein, O. (1920). Experimentelle Beiträge zur Lehre von den katatonischen Pupillenveränderungen. *Eur. Neurol.* 47, 194–215. doi: 10.1159/000190690

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *J. Cogn.* 1:16. doi: 10.5334/joc.18

Mayer, R. E. (2005). "A cognitive theory of multimedia learning," in *The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer (Cambridge: Cambridge University Press), 41–61.

Mayer, R. E. (2014). "Cognitive theory of multimedia learning," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer (Cambridge: Cambridge University Press), 43–71.

Mayer, R. E. (2021). *Multimedia Learning (Third Edition)*. Cambridge: Cambridge University Press.

Mayer, R. E., and Fiorella, L. (2014). "Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*. ed. R. E. Mayer (Cambridge: Cambridge University Press), 279–315.

Miller, A. L., and Unsworth, N. (2020). Variation in attention at encoding: insights from pupillometry and eye gaze fixations. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 2277–2294. doi: 10.1037/xlm0000797

Mitra, R., McNeal, K. S., and Bondell, H. D. (2017). Pupillary response to complex interdependent tasks: a cognitive-load theory perspective. *Behav. Res. Methods* 49, 1905–1919. doi: 10.3758/s13428-016-0833-y

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *ETRA*, 141–144. doi: 10.1145/1743666.1743701

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology* 11, 559–566. doi: 10.1111/j.1469-8986.1974.tb01114.x

Pomerleau-Turcotte, J., Moreno Sala, M. T., Dubé, F., and Vachon, F. (2021). Does cognitive load differ among sight-singers? An exploratory study using pupillometry and interviews. In *Proceedings of the Future Directions of Music Cognition International Conference March 6-7, Virtual Conference*. Columbus: Ohio State University Libraries, pp. 158–162.

R Core Team. R: a language and environment for statistical computing (version 4.0.4) [computer software]. (2021). *R Foundation for Statistical Computing*. Vienna, Austria. Available at: https://www.R-project.org/

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cogn. Sci.* 38, 1–37. doi: 10.1111/cogs.12086

Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–36. doi: 10.1016/j.edurev.2015.12.003

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2020). Decoding case comparisons in organic chemistry. Eye-tracking students' visual behavior. *J. Chem. Educ.* 97, 3530–3539. doi: 10.1021/acs.jchemed.0c00418

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2021). Connecting explanations to representations: benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry. *Int. J. Sci. Educ.* 43, 2707–2728. doi: 10.1080/09500693.2021.1985743

Rodemer, M., Lindner, M. A., Eckhard, J., Graulich, N., and Bernholt, S. (2022). Dynamic signals in instructional videos support students to navigate through complex representations: an eye-tracking study. *Appl. Cogn. Psychol.* 36, 852–863. doi: 10.1002/acp.3973

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., and Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instr. Sci.* 43, 93–114. doi: 10.1007/s11251-014-9328-3

Schneider, S., Beege, M., Nebel, S., and Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educ. Res. Rev.* 23, 1–24. doi: 10.1016/j.edurev.2017.11.001

Sharma, H., Drukker, L., Papageorghiou, A. T., and Alison Noble, J. (2021). Multi-modal learning from video, eye tracking, and Pupillometry for operator skill characterization in clinical fetal ultrasound. In *Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging April 13-16*, Nice, Piscataway: Institute of Electrical and Electronics Engineers, Inc., 1646–1649.

Shechter, A., and Share, D. L. (2021). Keeping an eye on effort: a Pupillometric investigation of effort and effortlessness in visual word recognition. *Psychol. Sci.* 32, 80–95. doi: 10.1177/0956797620958638

Stanners, R. F., Coulter, M., Sweet, A. W., and Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motiv. Emot.* 3, 319–340. doi: 10.1007/BF00994048

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive Load Theory*. Springer New York, NY.

Sweller, J., van Merriënboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., and Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Szulewski, A., Kelton, D., and Howes, D. (2017). Pupillometry as a tool to study expertise in medicine. *Frontline Learn. Res.* 5, 55–65. doi: 10.14786/flr.v5i3.256

Tabachnick, B. G., and Fidell, L. S. (2007). *Experimental Designs Using ANOVA*. Belmont, CA: Thomson/Brooks/Cole.

van der Wel, P., and van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: a review. *Psychon. Bull. Rev.* 25, 2005–2015. doi: 10.3758/s13423-018-1432-y

van Gog, T. (2014). "The signaling (or cueing) principle in multimedia learning," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning.* ed. R. E. Mayer (Cambridge: Cambridge University Press), 263–278.

van Merriënboer, J. J. G., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686–1691. doi: 10.21105/joss.01686

Xie, H., Wang, F., Hao, Y., Chen, J., An, J., Wang, Y., et al. (2017). The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: a meta-analysis and two meta-regression analyses. *PLoS One* 12:e0183884. doi: 10.1371/journal.pone.0183884

Zheng, R., and Cook, A. (2012). Solving complex problems: a convergent approach to cognitive load measurement. *Br. J. Educ. Technol.* 43, 233–246. doi: 10.1111/j.1467-8535.2010.01169.x

Zhou, C., Lorist, M. M., and Mathôt, S. (2022). Is categorization in visual working memory a way to reduce mental effort? A Pupillometry study. *Cogn. Sci.* 46:e13194. doi: 10.1111/cogs.13194

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/feduc.2023.1140272

OPEN ACCESS

EDITED BY
Pascal Klein,
University of Göttingen, Germany

REVIEWED BY
Sebastian Becker,
University of Cologne, Germany
Nicole Graulich,
University of Giessen, Germany

*CORRESPONDENCE
Matthias Schweinberger
m.schweinberger1@lmu.de

RECEIVED 08 January 2023
ACCEPTED 05 May 2023
PUBLISHED 30 May 2023

CITATION
Schweinberger M, Watzka B and Girwidz R (2023) Eye tracking as feedback tool in physics teacher education.
Front. Educ. 8:1140272.
doi: 10.3389/feduc.2023.1140272

Eye tracking as feedback tool in physics teacher education

Matthias Schweinberger[1]*, Bianca Watzka[2] and Raimund Girwidz[1]

[1]Faculty for Physics, Ludwig-Maximilians-Universität München, München, Germany, [2]Institute of Physics (IFP), Didactics of Physics, Otto-von-Guericke-Universität, Magdeburg, Germany

The ability to direct pupils' attention to relevant information during the experimental process is relevant for all science teachers. The aim of this article is to investigate the effects of training the ability of prospective physics teachers to direct attention during the presentation of experiments with eye tracking visualizations of pupils' visual attention as a feedback tool. Many eye tracking studies in the field of learning use eye movement recordings to investigate the effectiveness of an instructional design by varying cues or the presentation format. Another important line of research relates to study the teacher's gaze in a real classroom setting (mobile eye tracking). Here we use eye tracking in a new and innovative way: Eye tracking is used as a feedback tool for prospective teachers, showing them the effects of their verbal moderations when trying to direct their pupils' attention. The study is based on a mixed methods approach and is designed as a single factor quasi-experiment with pre-post measurement. Pre- and post-test are identical. Prospective teachers record their verbal moderations on a "silent" experimental video. The quality of the moderation is rated by several independent physics educators. In addition, pupils' eye movements while watching the videos are recorded using eye tracking. The resulting eye movements are used by the lecturer to give individual feedback to the prospective teachers, focusing on the ability to control attention in class. The effect of this eye tracking feedback on the prospective teachers is recorded in interviews. Between the pre-test and the post-test, the results show a significant improvement in the quality of the moderations of the videos. The results of the interviews show that the reason for this improvement is the perception of one's own impact on the pupils' attention through eye tracking feedback. The overall training program of moderating "silent videos" including eye tracking as a feedback tool allows for targeted training of the verbal guidance of the pupils' attention during the presentation of experiments.

KEYWORDS

silent videos, eye tracking, feedback, directing attention, self-perception, verbal cues

1. Introduction

1.1. Directing attention

1.1.1. Learning and the role of paying attention

The Cognitive Theory of Multimedia Learning (CTML) is based on three assumptions that are known as the cognitive principles of learning. The first principle is that multimedia learning takes place via a visual and an auditory processing channel. The second is that both channels have a limited capacity. Accordingly, learners can only process a certain amount of information per channel at the same time. The last is the principle of active learning. It states that learning takes place through cognitive processes (Mayer, 2014). Mayer (2014) identifies five cognitive processes: (1) selecting relevant words, (2) selecting relevant images, (3) organizing the selected words into a coherent verbal representation, (4) organizing the selected images into a coherent pictorial representation, and (5) integrating the pictorial and verbal representations and prior knowledge. The selection of words or images implies that learners pay attention to the information presented (Mayer, 2014). With regard to attention, a distinction is made between visual attention, auditory attention and other forms that are not relevant here (see Amso, 2016). With the help of "silent videos," we want to investigate the ability of prospective teachers to direct pupils' visual attention through speech. Therefore, only visual and auditory attentions are relevant.

- Visual Attention. Lockhofen and Mulert (2021) further specify the role of attention in the learning process. They define: "*Visual attention is the cognitive process that mediates the selection of important information from the environment.*" (Lockhofen and Mulert, 2021, p. 1).
- Auditory Attention. *"It is well-known that stimulus-focused attention improves auditory performance by enabling one to process relevant stimuli more efficiently"* (Folyi et al., 2012, p. 1).

Another distinction is the trigger that activates attention. Katsuki and Constantinidis (2014) distinguish between bottom-up attention and top-down attention. Bottom-up attention is an externally induced process. The information to be processed is selected automatically. Top-down attention is an internally generated process. The information is actively sought out based on self-selected factors (Katsuki and Constantinidis, 2014). The reasons for the attention diversion are different for bottom-up and top-down. However, their effects are similar. In both attentional processes, objects are processed preferentially. In both cases, a stronger neural response follows, which can induce better storage in memory (Pinto et al., 2013).

Therefore, both forms of attention may be of interest for teaching. The bottom-up process is a *stimulus-driven process* (Pinto et al., 2013). So, it could be specifically triggered by signals or cues to direct visual or auditory attention to relevant information. The top-down process is influenced by prior knowledge (Lingzhu, 2003; Lupyan, 2017) or previous experience (Addleman and Jiang, 2019).

1.1.2. Controlling attention through cueing

Cues are often defined as content-free information that is intended to direct attention and thus support cognitive processes (Hu and Zhang, 2021). Spotlights (de Koning et al., 2010; Jarodzka et al., 2013), color changes (e.g., Ozcelik et al., 2010) and arrows (Kriz and Hegarty, 2007; Boucheix and Lowe, 2010) are good ways of directing visual attention. However, cues often differ greatly, and not only in how they appear, when they appear, or what they look like. The classical categorization by modality (e.g., auditory or visual) does not do justice to this fact. One category that has received little attention so far is the question of the content richness of the cues (Watzka et al., 2021), which by definition should be absent. However, in many classic examples, such as the label, it is present. A label therefore has a different quality than a spotlight, which is only intended to direct attention. In this study, only verbal cues are used, which can be offered with or without content. For example, one can direct attention ("Look to the left!"), another can help with specific details ("the wooden block is an opaque object").

In meta-analyses regarding different subject areas, Richter et al. (2016), Schneider et al. (2018), and Alpizar et al. (2020) confirm the positive effect of the cueing principle on learning especially for novices. The analysis of Richter et al. (2016) includes 27 studies. Their main finding is that cues have a positive effect on learning performance with small to medium effect sizes and that especially learners with low prior knowledge benefit from cues. The analysis by Schneider et al. (2018) includes 103 studies and also includes eye tracking data. In summary, they also confirm the beneficial effect of cueing on learning success. In addition, attentional cues with small to medium effect sizes seem to induce longer learning times in general and longer gaze durations on relevant information in particular (Schneider et al., 2018). The mean gaze duration can be attributed to the cognitive process of organizing the CTML (Alemdag and Cagiltay, 2018) and indicate the degree of mental effort (Jarodzka et al., 2015). Ozcelik et al. (2010) interpret long mean gaze duration as more demanding tasks and correspondingly higher mental effort. Cues lead to longer viewing of the information addressed by the cues in learning materials than in learning materials without cues (Boucheix and Lowe, 2010; Ozcelik et al., 2010; Glaser and Schwan, 2015; Xie et al., 2019).

In a predominantly image-based learning material such as videos, verbal cues in particular have a positive effect on visual attention and learning success (e.g., Glaser and Schwan, 2015). An explanation for the better suitability of spoken text compared to written text can be found in CTML (Mayer, 2014). Due to the limited capacity of the processing channels, it makes sense to use additional resources of the auditory processing channel and thus follow the modality principle (see section "1.1.3. Modality principle and learning with experimentation videos").

1.1.3. Modality principle and learning with experimentation videos

The modality principle generally means that it is beneficial for learning if the text which accompanies graphics is spoken instead of written. Among other things, the modality principle has a positive effect on visual attention, because there is no split-attention effect to worry about (Schmidt-Weigand et al., 2010). In a predominantly image-based learning material (e.g., videos), spoken texts (e.g., moderations) in particular have a positive effect on visual attention and learning (Glaser and Schwan, 2015).

In a meta-analysis comprising of 43 studies which cover a vast spectrum of subjects and visualizations, Ginns (2005) confirms the modality effect with a medium effect size and shows that learning materials with visualizations and spoken texts generally lead to better learning outcomes than learning materials with visualizations and written texts.

The beneficial learning impact of the modality effect is explained by a more effective use of working memory capacity (see section "1.1.1. Learning and the role of paying attention"). Accordingly, more cognitive resources can be used for processing the learning content and learning performance increases (Sweller et al., 2011). When demonstrating experiments in class, teachers automatically use their voice as their main tool of communication. They automatically give verbal cues, some of which are content-related (e.g., mentioning the function) and some of which control attention (e.g., mentioning a surface feature). The question is how do prospective teachers learn to control the attention of their pupils? This paper is about fostering prospective physics teachers to guide their pupils in selecting relevant information by controlling bottom-up visual attention during experimentation through verbal cues. The control of visual bottom-up attention in this study is done via the cueing principle (verbal cues) since this technique can be applied without effort to classroom practice when teachers present experiments. Support for prospective teachers provides a special feedback format, which is theoretically classified subsequently.

1.2. Feedback

1.2.1. Definition and phases

"Feedback is information provided by an agent regarding aspects of one's performance or understanding" (Hattie and Timperley, 2007, p.81). Focusing especially on learners, Shute defines feedback as *"information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning"* (Shute, 2008, p.153). Feedback thus shows the gap between the target and the current state and should enable the recipient to recognize and close this gap. In this study, the presentation of pupils' gaze behavior is intended to provide feedback and to help prospective teachers become aware of their ability to control attention. The three classic feedback phases described in the literature occur, namely, (Wisniewski et al., 2020):

- "Feed-up" (comparison of the actual status with a target status) Students and teachers get information about the learning goals to be accomplished: By watching the gaze overlays of their first moderation (pre) the prospective teachers got information about how the pupils reacted to their moderation of the video.
- "Feed-back" (comparison of the actual state with a previous state) Students and teachers see, what they have achieved in relation to an expected standard or previous performance: By watching at the gaze overlay of their second try, the prospective teachers could see what they have achieved relative to their first performance.
- "Feed-forward" (explanation of the target state based on the actual state), Students, and teachers receive information that leads to an adaption of learning in the form of enhanced challenges: After analyzing both moderations, the prospective teachers became aware of the positive skills they should develop, and the mistakes they should avoid in the future.

In general, feedback is considered a very powerful tool. Wisniewski et al. (2020) obtain an average effect size of 0.48 in a meta-analysis. However, feedback does not *per se* lead to better learning outcomes. Kluger and DeNisi (1996) note that about one third of feedback results in negative learning effects. However, learning depends on a variety of different influences (Hattie, 2021), so there is no standardized way to use feedback. What helps one student today may not help another. Tomorrow, the same feedback may have the opposite effect or no effect at all (Hattie, 2021). How feedback is received depends not only on the form in which it is given, but also on a variety of factors about the recipient (Shute, 2008). For example, important factors are the recipient's self-assessment and experience of self-efficacy (Shute, 2008).

1.2.2. Levels and forms of feedback

To understand the effectiveness of feedback, one must first be aware of the different levels that feedback addresses (Hattie and Timperley, 2007). Firstly, feedback works at the task level (FT). Is the answer on the task wrong, or right? Second, feedback addresses the process level (FP), i.e., information about the process, how to deal with the task and/or how to understand it. Thirdly, feedback works on the self-regulation level (FR), where the learner checks, controls, and self-regulates his or her processes and behavior. Finally, feedback also provides feedback on the so-called self-level (FS), where positive (and negative) expressions and evaluations about the learner are expressed (Hattie and Timperley, 2007). Eye tracking feedback on your own moderation should ideally trigger the task and process level.

The level of feedback addressed depends largely on the form in which it is given. Different authors distinguish between written, computer aided, oral, pictorial, etc., according to the medium, or according to the content, for example, formative tutorial (Narciss and Huth, 2006) or actionable (Cannon and Witherspoon, 2005). A detailed description of the different forms can be found in Hattie and Timperley (2007) and Wisniewski et al. (2020). By watching the gaze overlays of individual moderated videos, we concentrate on a certain form of visual and auditive feedback (see section "1.2.5. Eye tracking as feedback tool" and section "3.2.1. Pre-test, first eye tracking feedback and pre-interview").

1.2.3. Feedback directions/student feedback

Much of the research describes forms and effects of feedback on the learner. Recently, feedback as feedback from the learner to the teacher has received more attention (Rollett et al., 2021). The focus here is on the question of the extent to which pupil feedback affects the quality of the teacher's teaching and thus improves the pupil's learning success. The question is to what extent pupil feedback is reliable and valid, but recent studies show that pupil feedback provides teachers with valid information about their teaching quality (Rollett et al., 2021). In this study, training with pupils' gaze overlays should provide valid information for feedback, especially since the pupils provide this feedback without their own knowledge.

Röhl et al. (2021) describe in the "Process Model of Student Feedback on Teaching (SFT)" a circuit diagram of how pupil feedback affects the teacher. The process begins with collecting and measuring pupil perceptions, which are then reported back to the teacher. The teacher interprets this feedback information, which stimulates cognitive but also affective reactions and processes in the teacher. This information can increase the teacher's knowledge about his teaching and thus trigger a development to improve his own teaching, so that in the following the learning success of the pupils can increase again. By giving the prospective teachers feedback information about their moderations we assume that the development will be triggered to better direct the attention of pupils.

1.2.4. Eye tracking as feedback tool

The use of eye tracking as a feedback tool in education has recently been increasingly emphasized in various disciplines (e.g., Cullipher et al., 2018). Eye movement recordings have been used to analyze and optimize the effectiveness of the design of learning materials (Langner et al., 2022). Mussgnug et al. (2014) describe how eye tracking recordings as a teaching tool improve awareness of user experiences with designed objects and how these experiences can be implemented in design education. Xenos and Rigou (2019) outline the use of eye tracking data collected and analyzed to help students improve their design. In contrast to gaze data of other people looking at specific objects, the gaze of teachers in real classrooms has also been the subject of various studies (McIntyre et al., 2017; Stuermer et al., 2017; McIntyre and Foulsham, 2018; Minarikova et al., 2021). In addition to using the gaze data of others, one's own gaze can also be used as feedback (Hansen et al., 2019). Szulewski et al. (2019) investigated the effect of eye tracking feedback on emergency physicians during a simulated response exercise, presumably triggering self-reflection processes. Keller et al. (2022) examined the effect of eye tracking feedback on prospective teachers observing and commenting on their own gaze during a lesson they were teaching.

We use eye tracking in a different way, somewhere in between the above: Eye tracking is used as a feedback tool for prospective teachers, showing them the effects of their verbal moderations as they try to direct their pupils' attention, as happens in the regular classroom.

2. Research question

Directing pupils' attention during the presentation of an experiment is crucial to its success. Pupils need to look at the right time at the right place to make the important observations. External cues such as speech can influence visual attention (Glaser and Schwan, 2015; Xie et al., 2019; Watzka et al., 2021). The overall question is how to improve the competence of prospective teachers in moderating experiments in the classroom. Therefore, we used the method of moderating "silent videos" to train prospective teachers' ability to control their pupils' attention. The particular focus of this method is on verbal cues through spoken language during the presentation of a video. Based on the five cognitive processes (Mayer, 2014), one of the main objectives of an appropriate presentation is to allow pupils to make the necessary observations, among many other aspects (see section "3.3.1. Assessment of prospective teachers' competence in moderating experimental videos"). To assess this process, the times when observation tasks are set and when pupils are explicitly given the opportunity to observe are summarized as "pupil-activating time".

Eye tracking is often used to study how a stimulus affects a person's perception. Conversely, visualizations of eye tracking data can be used to draw conclusions about the observer's attention and the effectiveness of cues. By using eye tracking as a feedback tool, we tried to show prospective teachers the impact of their moderation of an experimental video on pupils, so that they in turn could draw consequences for further presentations. This leads to the following research questions.

> *RQ: To what extent can training with eye-tracking visualizations of pupils' visual attention improve prospective teachers' guidance of pupils' gaze?* The following more detailed questions should be considered.
> *RQ1: Does training with eye tracking feedback help prospective teachers explain the set-up of an experiment in a way that is adapted to pupils' prior knowledge and cognitive and linguistic development?*
> *RQ2: Does training with eye tracking feedback help prospective teachers to increase pupils' activating time?*

3. Materials and methods

3.1. Participants

A subsample of 15 physics prospective teachers from a German (Bavarian) university was selected. They were on average 22.4 years old (SD = 3.4) and in the 5th semester. Of the participants two were female and 13 were male. All participants had heard the experimental physics lectures and an introductory lecture on physics education with a theoretical introduction to criteria for setting up and conducting experiments before the study. Thus, all students had the necessary content and pedagogical knowledge on the topic of the study.

3.2. Procedure and material

The study uses a pre/post-test design. Between the pre-test and the post-test, a training phase of several weeks took place for the moderation of demonstration experiments. "Silent videos" were used both in the pre- and post-test as a survey instrument and in the training as learning material. The overall process of the study is shown in **Figure 1**.

3.2.1. Pre-test, first eye tracking feedback and pre-interview

As a pre-test the prospective teachers had to moderate a "silent video" about core shadows and semi shadows (see **Figure 2**). The video is divided into two main parts: a static part showing the set-up for about 30 s and a dynamic part showing the execution of the experiment for about 60 s. The video shows a small opaque block that is illuminated by two sources from different angles. A white elongated rail serves as a screen on which the different kinds of shadow can be seen. Everything is recorded from the pupils' perspective, and it is presented in real time. All activities are shown as they would normally be done in a live classroom demonstration.

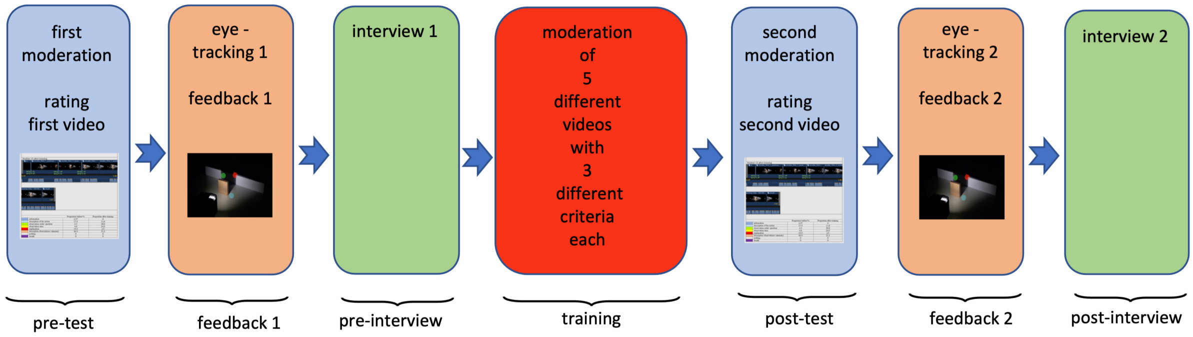


FIGURE 1
Procedure of the entire study.

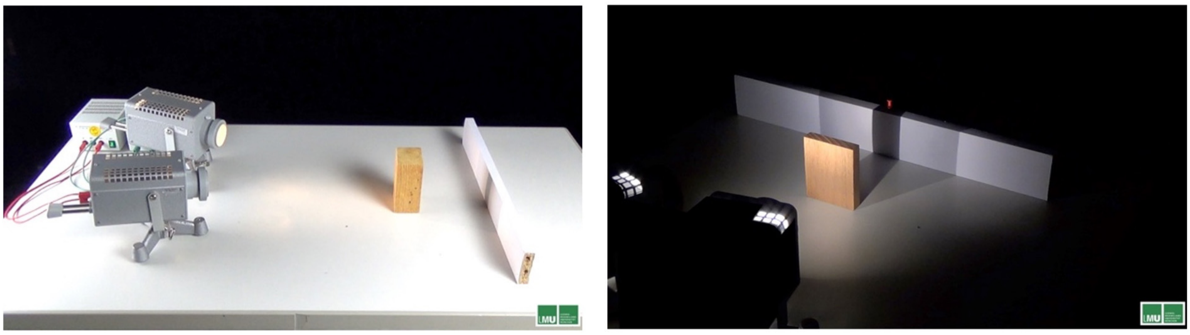


FIGURE 2
Screenshot of set-up and execution of the study's experiment. The method of "silent videos" is described in detail by Schweinberger and Girwidz (2022).

For further information[1] about the training with "silent videos" (see Schweinberger and Girwidz, 2022).

The task for the prospective teachers was to moderate the video appropriately for pupils in their first year of learning physics in junior high school. The prospective teachers were told to assume that their pupils had prior knowledge of the model of the rectilinear propagation of light and the appearance of cast shadows. The moderation of the videos was evaluated by four or five raters according to the criteria (see section "3.3.1. Assessment of prospective teachers' competence in moderating experimental videos").

In the next step, individual feedback was given to the prospective teachers. For the eye tracking feedback, each video moderated by the prospective teachers was shown to three randomly selected pupils of the 7th grade and their eye movements were recorded using eye tracking. The data from this tracking was used to create a single gaze overlay video, in which the three gaze overlays of the pupils were superimposed (see **Figure 3**). When the three overlays are superimposed at the same time, it is easier to see the commonality of the pupils' responses than when all three overlays are viewed in sequence.

The feedback took the form of a short, written critique by the lecturer before the discussion of the gaze overlay video. The gaze overlays were used to illustrate to the prospective teacher the immediate consequences of the criticisms previously made. In an individual conversation the lecturer and every prospective teacher watched the video with the gaze overlays together and discussed the connection between moderation and the pupils' reactions (e.g., did the pupils look to the area they should and how long did they stay). For this purpose, the gaze overlay feedback was stopped at important points to study certain situations more intensively. If necessary, the gaze overlay feedback was viewed several times.

Afterward the prospective teachers were interviewed for the first time by another research assistant. They were asked to what extent the feedback from the pupils' views helped them to assess their own ability to manage their pupils' attention (see section in details "3.3.2. Interview and survey guide").

3.2.2. Training phase

The prospective teachers moderated a total of six videos over a 10-week period later in the term to build up their skills. Like the pre-test, in the training phase the prospective teachers had to moderate "silent videos" of different experiments. To do this, they had to write their own script in advance, considering the criteria. In order to train as many facets of a moderation as possible, three criteria from the catalog (see list of criteria in the **Supplementary Figure 1**) were given by the lecturer for each training video. The main focus was on the development of the

1 For all "silent videos" see: https://www.didaktik.physik.uni-muenchen.de/lehrerbildung/lehrerbildung_lmu/video/

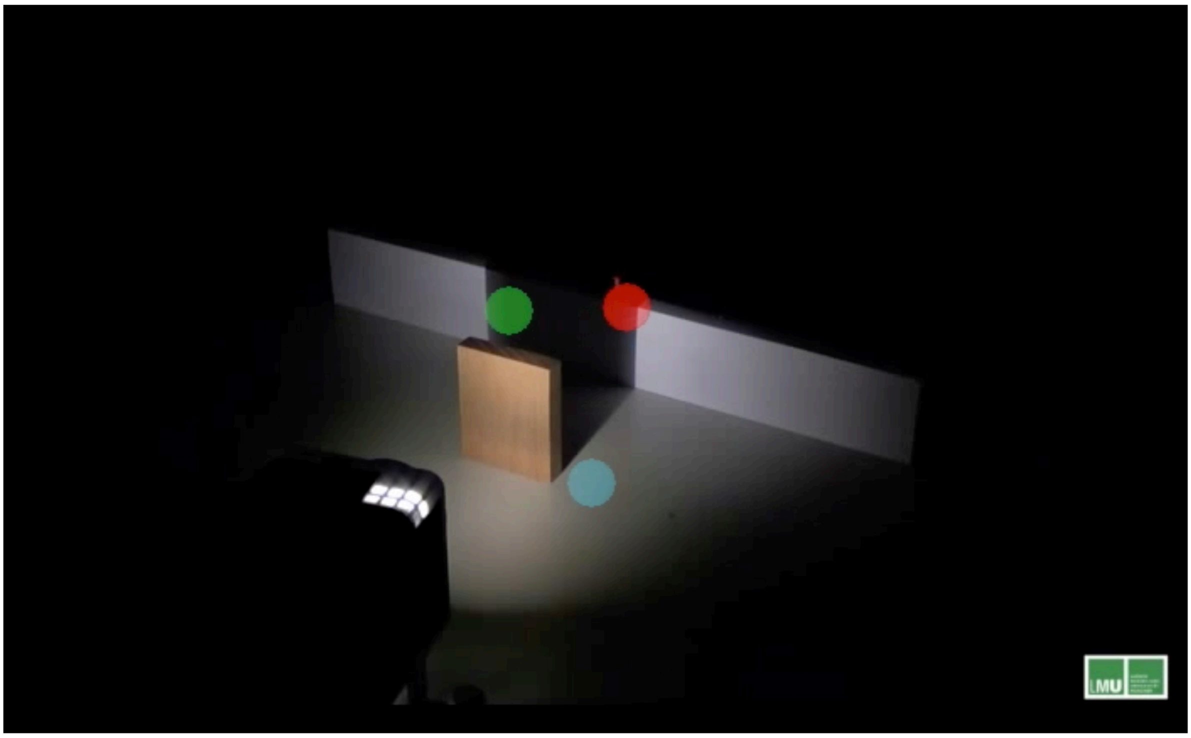


FIGURE 3
Video excerpt of the gaze overlay feedback to a prospective teacher's moderation of the shadow video. The colored dots are the gaze overlays of three different pupils watching and reacting one prospective teacher's moderation.

prospective teachers' ability to direct pupils' attention in a targeted way. Each of these moderations was analyzed individually in a small group discussion based on the three pre-defined criteria. In preparation for these discussion meetings, each prospective teacher received a brief written critique in advance. Afterward, they had to set up the respective experiment in the seminar and present it to their colleagues. Thus, the prospective teachers received verbal and written feedback from the lecturer and their student colleagues in the training phase, no further gaze overlays were shown (see section "6. Limitations").

3.2.3. Post-test, second eye tracking feedback and post-interview

After the training, at the end of the term, the prospective teachers had to moderate the first video about core shadows and semi shadows from the pre-test for a second time (post-test). The moderation of the videos in the post-test was also evaluated according to the criteria (see section "3.3.1. Assessment of prospective teachers' competence in moderating experimental videos") as in the pre-test.

To generate the second eye tracking feedback, the moderated videos from the post-test were again shown to three different pupils and their gazes were recorded. We decided to use different pupils than in the pre-test, because we expected quite a large repetition effect. The content was a very simple phenomenon, and we wanted all pupils to have the comparable prior knowledge. The resulting gaze overlay videos were produced as in the first feedback and shown to the prospective teachers visualizing a second short written critique. In addition, the prospective teachers watched the gaze overlay video of their first trial, to discuss the developments between the pre and post.

The prospective teachers were then interviewed a second time by another research assistant using the same questions as in the first interview.

3.3. Assessment

3.3.1. Assessment of prospective teachers' competence in moderating experimental videos

The criteria were developed over several years from practical experience and then discussed intensively by five physics lecturers from the chair of Physics Education at LMU Munich and two physics teacher trainers. The criteria are subject to constant further development. Due to the two different parts of the video (static set-up and dynamic execution), two evaluation schemes had to be developed (which also were explained to the prospective teachers).

In the set-up, each relevant object had to be described by three categories of the object in the experiment:

- the location (e.g., "on the left side of the table"),
- two surface characteristics (e.g., "brown, wooden block"), and
- the function (e.g., "provides shade").

Reading from left to right results in three consecutive sequences: first the lamps, then the block, and finally the screen. For each of the relevant objects, the number of mentions was counted.

General information	The content of the moderation is not directly related to the experiment.
Description of action carried out	The moderation describes the action being performed at the same time as the action shown in the video.
Observation tasks / questions	The moderation involves giving tasks or asking questions that lead the students to observe certain aspects.
Observation time	The phases during the moderation, without any comment from the prospective teacher, that allow the pupils to make observations on their own.
Explanations	The phases during the moderation, which include explanations from the prospective teachers on the intended observations to the pupils.
Summary of observations	Time during the moderation used to summarize what was observed.
Silent time	The phases during the moderation, in which the prospective teachers do not speak, but the pupils are also left without an observation task or question.
Breathing time	The phases of deep breathing during the moderation. But the moderator was in action and didn't leave the pupils without a task or a question.

FIGURE 4
Coding scheme for moderating the execution of the experiment. The colors represent the categories.

Different categories were chosen to assess the moderation of the execution of the experiments (see Figure 4):

The following procedure was used to assess the extent to which the moderations met the criteria. After adding the prospective teachers' audio tracks to the "silent videos," the different categories were localized in the timeline and marked by a corresponding, colored bar. The categories were: general information to the experiment (blue), description of action carried out (light blue), observation orders or questions (green), observation time (yellow), explanations (red), summary of the observations (orange) and time without content (no color), breathing time (purple). The length of the bars is proportional to the temporal length of these intervals. The codes are rated on whether they appear. Then the ratios of the corresponding intervals to the total length of the moderation of the execution were calculated. This was done for both moderations of pre and post-test (see Figure 5).

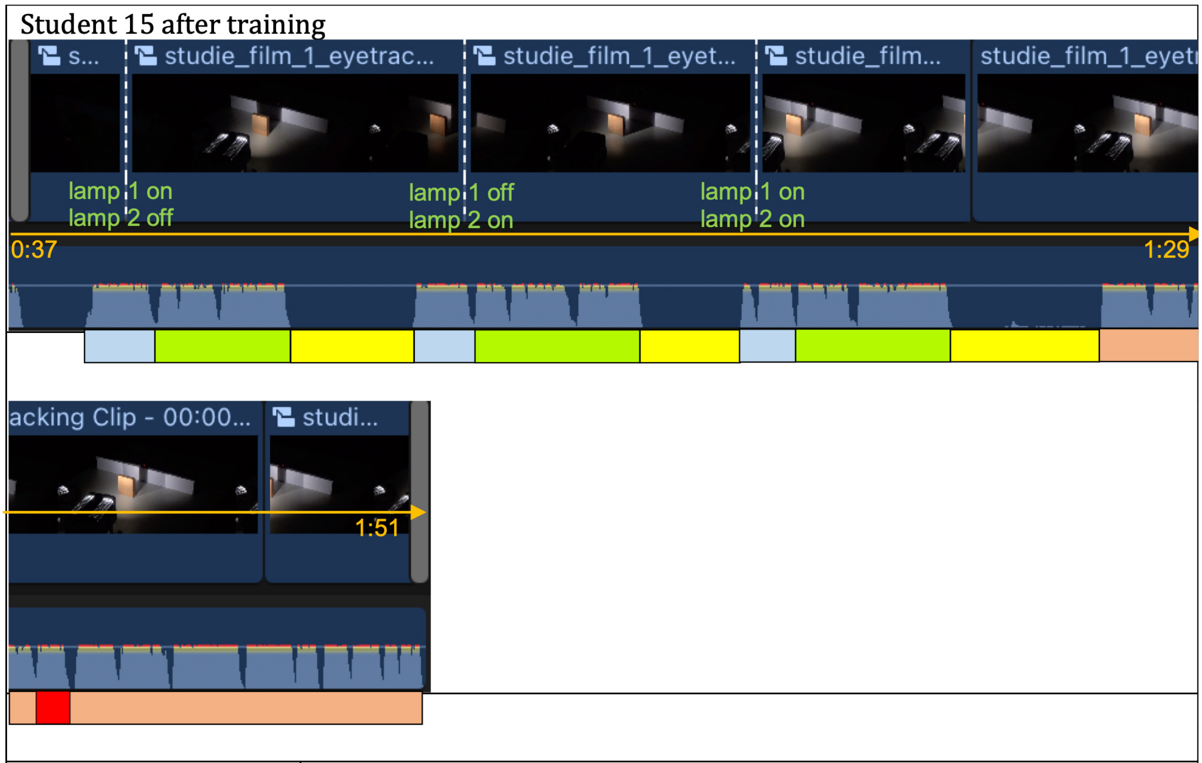
This, of course, raises the question of the relationship of each category to each other for optimal moderation. Discussions with various training teachers and lecturers led to the conclusion that it is difficult, if not impossible, to define an ideal moderation of an experiment. Different individual teaching styles and current classroom situations are too different. We limited ourselves to the general consensus that a good moderation must give the pupils the opportunity to make the necessary observations, i.e., the pupils must be activated to do these observations. Also, a good moderation should also not involve explanations, as these disrupt the observation process, deprive pupils of the opportunity to think through the process themselves, or become the content of subsequent lessons. To assess these pupil-activating segments, we added both observation time and observation orders to the one total pupil-activating time.

3.3.2. Interview and survey guide

The interview and survey guide included 13 questions, the eight questions used for this research were the same in pre-and post-interviews. The prospective teachers' ratings were recorded using single items in the form of a 4-point Likert scale (4: completely agree," "3: agree," "2: disagree," and "1: completely disagree). To obtain detailed information about their specific experiences, an open-ended question about this item was added. The interview and survey guide contained questions about the effects of moderating "silent videos" and of getting eye tracking feedback on their personal learning process (see Supplementary Figure 2). They should describe how their skills in controlling attention in particular and in moderating the videos in general had changed (Q 2). They were also asked about the effects on their professional language (Q 3, 6) and the consequences for their own actions in experimentation (Q 7). Another important part of the interview questions was the prospective teachers' experiences of eye tracking as a feedback tool. They were asked how they perceived the effectiveness of their facilitation on the pupils. A major question was how eye tracking showed the connection between guiding (tasks and questions) and the pupils' attentional response (Q 9, 10, 11, and 12). Finally, the prospective teachers were asked how they rate their learning progress between the two measurement points concerning approach, controlling attention through using language in facilitating experiments (additional question in the second interview). All interviews were evaluated and analyzed by two independent persons.

3.4. Eye tracking system

In this study, eye tracking was used as a feedback tool. It is therefore not a measurement tool to measure an outcome variable, instead it is a part of the intervention/training. The eye movements were recorded with an eye tracker. The system used was an Eye Follower from LC Technology. This system uses four cameras, two for tracking head motions, and two for tracking the eyes. The accuracy was less than 0.4° of visual angle. The distance of a participant to the monitor was between 55 and 65 cm. The video area has a resolution of 1920 × 1080 pixels and the resolution of the 24″ monitor is 1920 × 1200 pixels. The stimulus was enlarged to full monitor width and proportionally adjusted in height. The fixations and saccades were recorded at a sampling rate of 120 Hz and the discrimination between saccades and fixations was done by LC Fixation Detector (a dispersion-based algorithm: Salvucci and Goldberg, 2000).



	transcript
information	None given
description action	*"… Now the spotlight at the upper edge of the image is switched on…"*
observation order	*"… Observe how the lighting conditions behave on the bar. Is there any difference from..?"*
observation time	
explanation	*"… shine both spotlights on…"*
summary	*"…if one spotlight is switched on, we can see a shadow on the bar…"*
nothing	

FIGURE 5
Timeline with rating of a prospective teacher's moderation: at the top the video track, below it the audio track and again below sample codes for the assignment to the categories.

3.5. Analysis

The moderations of the videos were rated by four to five independent raters based on the categories (see section "3.3.1.Assessment of prospective teachers' competence in moderating experimental videos"). The raters marked the beginning and end of each category on the timeline of the videos and calculated the percentage of time. The interclass reliability coefficient (model: two-way mixed and type: absolute agreement) was used to determine the agreement of the raters.

Dependent samples *t*-tests were used to test whether the mean speaking times per category differed between the pre-test and the post-test. The Bonferroni correction was used to counteract the accumulation of alpha errors by performing each individual test at a reduced significance level. The significance level of individual tests is calculated as the global significance level to be maintained divided by the number of individual tests (4 tests, significance level $\alpha = 0.0125$).

The interviews were analyzed using qualitative content analysis according to Mayring (2015). We followed a descriptive approach, analyzing the texts with a deductively formulated category system. We recorded the occurrence of these categories in category frequencies. The resulting scale has an ordinal scale level, so the "Cohen's Weighted Kappa" coefficient was calculated for the raters' agreement. We chose quadratic weights, where the distances between the raters are squared. This gives more weight to ratings that are far apart than to ratings that are close together.

4. Results

The moderation of the set-up was evaluated by four independent raters. The results of the rater agreement analyses show an agreement between the four raters of $r = 0.799$ [95% CI (0.686, 0.887)].

The moderation of the execution was evaluated by five independent raters. The value of the inter-rater correlation coefficient $r = 0.993$ shows a very high level of agreement between the five raters [95% CI (0.991, 0.994); see Cicchetti, 1994].

TABLE 1 Percentage of items "introductory sentence", "hypothesis mentioned" and "reading direction adhered" mentioned in the pre- and post-test.

	Pre	Post
Introductory sentence	43.3%	55.0%
Hypothesis mentioned	10.0%	22.7%
Reading direction adhered	93.3%	86.6%

The interviews were evaluated by two independent raters. The results of the rater agreement analyses show an agreement between the two raters of $\kappa = 0.694$ and are just above the 5% significance level [$\alpha = 0.068$; 95% CI (0.378, 1.010)].

The findings to answer the first research question, namely, whether eye tracking feedback helps prospective teachers to explain the experimental set-up in a way that is appropriate for pupils, are divided into a general and a specific part.

4.1. Set-up: general results

Before looking at the individual objects of the set-up to answer RQ 1, we will examine the connection of the set-up with the previous knowledge and the subsequent execution. A total of 43% of the prospective teachers started their first moderation attempt with an introductory sentence about the topic of the upcoming experiment, with only two of them really connecting to the pupils' prior knowledge. The number of participants who started with a reasonable introductory sentence increased to 55% in the post attempt. The number of participants moving from set-up to execution with a research question or hypothesis increased from 10 to 23%. In both cases, the low percentage indicates that the participants were not aware or did not become aware of the importance of the transition between set-up and execution of the experiment. With 93%, the overwhelming majority adhered to the reading direction (from left to right), with virtually all participants (except one) adhering to the reading direction in the post-trial when trying to direct the pupils' attention (see Table 1).

4.2. Set-up: specific results

Since the introductory sentence, the link to prior knowledge and the reading direction do not directly influence the pupils' visual attention to certain areas. Thus, there is no focusing effect on the observed gaze overlays; the pupils' gazes move across the whole screen and become more focused as soon as the experimental set-up appears, and the prospective teachers start talking. This behavior of the pupils didn't change between the pre- and post-trial.

After the introductory sentence the numbers of mentions regarding an object are counted (e.g., location, function and two surface features, see section "3.3.1. Assessment of prospective teachers' competence in moderating experimental videos"). The mentions for the lamps increased from 56 to 76% ($t = -4.636$, $p < 0.001$, Cohens' $|d| = 4.575$, $n = 15$), those concerning the block from 50 to 71% ($t = -15.756$, $p < 0.001$, Cohens' $|d| = 2.926$, $n = 15$) and the screen from 71 to 84% ($t = -9.1454$, $p < 0.001$, Cohens' $|d| = 1.698$, $n = 15$). The number of mentions increased for all subjects (see Figure 6).

4.2.1. Mentioning "block" (detailed)

A more detailed analysis–here of the description of the opaque block in the light path—provides further insights: Cues referring to the location of the block increased from 42 to 75% applicable mentions, while the description of the block's function in the experiment remained at about 33% (two participants who had mentioned the block's function in the first attempt didn't mention it in the second attempt.) Altogether, the function of the block seems to be too obvious for many prospective teachers to mention. In the post-attempt, all prospective teachers described the block with at least one surface feature, with the number of mentions increasing from 97 to 100%. A total of 77% of them

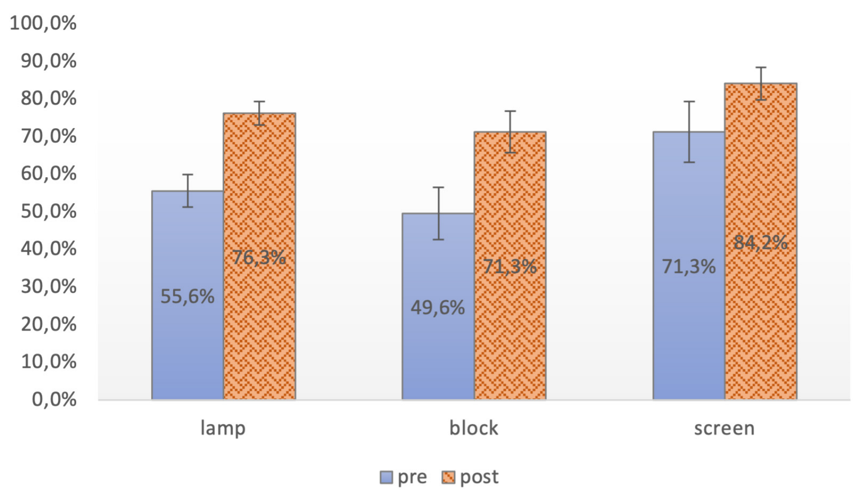


FIGURE 6
Percentage of the objects "lamps, block and screen" mentioned in the pre- and post-test.

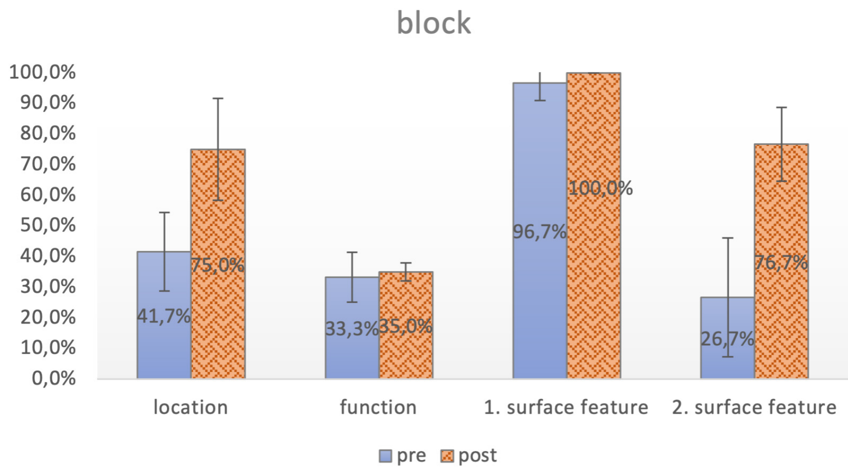


FIGURE 7
Percentage of the object "block" mentioned in detail in the pre- and post-test.

mentioned also a second feature, up from 27% in the first trial (see Figure 7).

4.3. Execution: specific results

A total of 60% of the prospective teachers did not ask any question or gave any observation order to the pupils in the first attempt. The same situation resulted for giving the pupils' time for observations, where also 60% of the prospective teachers didn't leave any time to do so. In the second moderation–after the training phase and the eye tracking feedback- all prospective teachers gave observation orders and time to do these tasks.

4.3.1. Activating time

Observation order time and observation time together result in the activating time. A total of 40% of the prospective teachers placed observation orders. The average order time increased from 2.5 to 13.7 s, which means an increase of time share from 4.3 to 18.9%. The same development can be seen in the amount of observation time given to the pupils. A total of 40% of the prospective teachers gave the pupils time for observations. The average observation time for all participants increased from 2.8 to 13.9 s, which means an increase of time share from 4.2 to 18.5%. Due to the high number of prospective teachers who did not give observation orders or time in the pre-trial, the *SD* is very high, so that the variance in response behavior is also large. The time span between prospective teachers activating pupils and non-activating is very large.

If we restrict ourselves to the participants who gave both observation order and observation time ($n = 7$), the following picture emerges:

The results of the dependent samples *t*-test show a significant difference with a high effect size between the mean percentage of activating time before and after moderation training with feedback [$t = -3.075$, $p = 0.033$, 95% CI (-29.21, -7.61), Cohen's $|d| = 15.77$, $n = 7$]. After training with eye tracking feedback ($M = 29.66\%$, $SD = 11.29$), subjects used significantly more pupil activating "tools" in their moderation than before training ($M = 11.33\%$, $SD = 11.13$). Due to the small sample size, a bootstrapping procedure with 10.000 samples was applied.

TABLE 2 Prospective teachers' time share and average time (pre and post) for observation orders, observation time and activating time when moderating the execution of an experiment.

	Time share pre	Time share post	Average time pre	Average time post
Observation orders	4.3%	18.9%	2.5 s	13.7 s
Observation time	4.2%	18.5%	2.8 s	13.9 s
Activating time	8.5%	37.4%	5.3 s	27.6 s

For prospective teachers who gave both orders and observation time in the first trial, the average length of orders increased from 6.5 to 15.3 s while the observation time given increased from 6.8 to 14.5 s. The share of pupil activating time more than doubled after the training (see Table 2 and Figure 8).

Overall, however, one of the most important findings is that all prospective teachers, regardless of what they did on the pre-trail, gave observation orders and observation time after the training with eye tracking feedback. The average percentage of pupils-activating time on the second trial was 34%, so that more than one-third of the execution time was used to activate the pupils.

4.3.2. Decrease of other parameters

Part of the training concept is that no explanations should be given while conducting an experiment. Explanations are major part of the next step in the lesson, only observations should be made and recorded during the experiment. Nevertheless, the percentage of explanations given by the prospective teachers decreased from 10% of time to 8.3% of time, with the number of teachers giving explanations remaining the

FIGURE 8
Increase of observation orders and observation time (pupils activating time) for prospective teachers who have given both in the first trial.

same. The prospective teachers did not seem to find this instruction meaningful.

The time spent on summaries also decreased from 42 to 30%, which seems to be a consequence of the fact that the prospective teachers were able to describe the essential content more precisely.

The descriptions of action followed the same trend as the summaries, falling from 19 to 15% of time, although the number of individual descriptions increased: The execution of the experiment is divided into three sequences (lamp1, lamp2, lamp1 and 2). In the post-trial 14 out of 15 prospective teachers gave concise and accurate action descriptions of these sequences. This was followed by observation tasks, with the timing of these messages much better aligned with the temporal sequence. The shorter duration of the action descriptions is again a consequence of the much more precise formulation of the descriptions.

Fortunately, the number of prospective teachers who temporarily left their pupils without any task dropped from eleven to five, i.e., by more than halved. The portion of time fell from 8.9 to 1.3% of the time, with only five instead of 11 prospective teachers leaving pupils without any instruction at all. All increases and decreases are shown in the following Sankey diagram (see Figure 9).

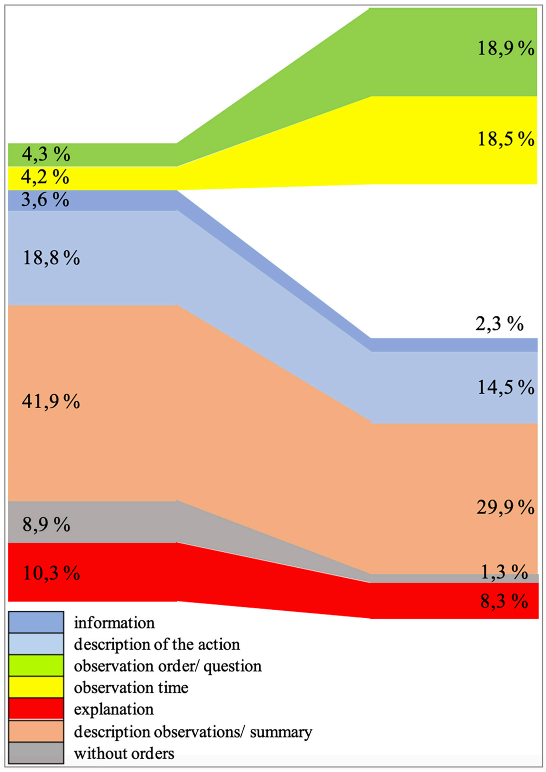


FIGURE 9
Sankey-diagram of all changes of the moderation of the execution of an experiment (all prospective teachers).

4.4. Results of the interviews and surveys

To answer the research question 2 (RQ2) of whether training with eye tracking feedback helps the prospective teachers to increase their pupils' activating time, we analyzed the interviews and surveys. We considered the following statements to evaluate and rate the interviews:

- Category 1: Awareness of own impact *"Eye tracking made me aware of my own impact on pupils."*

The Likert scale with the question "*Eye tracking made me aware of my own impact on pupils*" was answered in the pre-test with a mean M = 3.20 and SD = 0.75 and in the post-test with a mean M = 3.40 and SD = 0.49. This indicates that a large share of the prospective teachers showed high agreement with the statement and that this agreement even increased in the post-test. The decrease in SD shows that they even more agreed.

To evaluate the interviews regarding this category the following two key phrases were used: (1) "see reaction of the pupils." and (2) "see importance of orders."

Analyzing the interviews 63% of the participants fell into this category after the pre-trial, 70% after the post-trial. A comment of a prospective teacher (student_14) was: *"Eye tracking feedback is really good because you can just see how you're affecting the pupils. You're really doing something practical where you can directly see the consequences of your actions."*

- Category 2: Connection between control codes and pupils' reaction "*Eye tracking made me realize the connection between control codes (such as assignments and questions) and the response of the pupils.*"

The question "*Eye tracking made me realize the connection between control codes (such as assignments and questions) and the response of the students*" was answered in the pre-test with a mean M = 3.20 again and SD = 0.65 and in the post-test with a mean M = 3.73 and SD = 0.44. This indicates that a large share of the prospective teachers showed high agreement with the statement and that this agreement even increased in the post-test to very high agreement. The decrease in SD shows that they even more agreed. All prospective teachers of the study saw the pupils' reaction on the control codes they applied.

To evaluate the interviews regarding this category the following three key phrases were used: (1) "see the effect of the control codes," (2) "see effect of the spoken word," (3) "see where the pupils look to."

A total of 53% of the participants fell into this category after the pre-trial, 63% after the post-trial. "*You can clearly see where the children look during the experiment, especially how they react to instructions*", was one of the prospective teachers' comments (student_10).

- Category 3: Perceived difficulty in directing attention *"I found it easy to direct the attention of the pupils in a certain area of the experiment)."*

The Likert scale with the question "*Eye tracking made me aware of my own impact on pupils*" was answered in the pre-test with a mean M = 2.20 and SD = 0.65. After the first interview it became clear that the prospective teachers were rather reserved about their ability to direct pupils' attention. With a mean M = 2.87 and SD = 0.44 in the post-test it is obvious that the difficulties in directing the attention of the pupils decreased and the prospective teachers were in consensus about this development. However, agreement with this category lagged behind the others in all Likert-scored questions.

Responses that prospective teachers felt were important in directing pupils' attention were used to evaluate this category, i.e., the following three key words were used: (1) "location items," (2) ". . . surface features," (3) "observation order."

According to the interviews 63% of the participants fell into this after the pre-trial, 80% after the post-trial. *"I have noticed which work orders help more,"* commented student_9.

Another interesting finding from the interviews with the prospective teachers and the examination of many moderated videos must be mentioned. A total of 75% of prospective teachers indicated that it is not only important to keep the pupils' eyes on a particular area, but also to keep them there. So, it seems to be necessary to give the pupils after the order where to look at a second assignment so that the pupils' gazes stay on that spot.

This result was not expected in this way and was not previously part of our considerations of attention-controlling moderation of experiments. Rather, it seems necessary to investigate this circumstance more closely.

Generally, the overwhelming majority of prospective teachers stated, that *"it was very enlightening and informative to see where the pupils were looking and how they reacted to the instructions"* (student_13).

5. Discussion

In this study, we investigated a training with eye tracking feedback to improve prospective teachers' abilities of moderating experiments in class. To do this, we demonstrated to prospective teachers their ability to direct pupils' attention using only verbal cues. We encouraged these skills through training and intensive feedback on their abilities. The three phases of feedback described in the literature (Wisniewski et al., 2020) could be realized in our approach: "feed-up" was realized by the prospective teachers watching the gaze overlay videos of their moderation, "feed-back" by comparing pre- and post-trial, and "feed-forward" by becoming aware which skills they should develop and which they should avoid. This feedback consisted of an assessment of the quality of the moderation (rating) and, in particular, the pupils' reactions to the moderation (eye tracking). The success of this approach was measured by an assessment and through interviews.

Our results answering RQ 1 show that our approach significantly improves the ability to moderate experimental set-up through verbal cues. We could show that training with eye tracking feedback and rating feedback helped prospective teachers to explain the set-up of an experiment in a pupil's appropriate way. The number of categories mentioned by the prospective teachers increased for all three objects in the set-up. The prospective teachers rarely provided a second surface feature and often relied on the pupils' presumed prior knowledge. It is interesting to note that although the prospective teachers are better at locating the individual objects and at the same time name a second surface feature much more frequently, there is no significant change in all three objects in terms of their function. The function of an object also hardly appears in the interviews. To the prospective teachers the function of an object seemed to be automatically supplied with the naming of the object or not worth naming it. The wooden block seemed to be of little concern in both trials, although compared to lamps and screens, the block's function as a shade provider is not natural. When asked why the block was given so little attention, reference was made to the corresponding preliminary experiment, although only two prospective teachers and then only in the second trail made a sufficient connection to the previous knowledge (in this case the creation of a simple shadow in the model of the rectilinear propagation of light). Overall, the prospective teachers had difficulty making transitions between the different phases of the experiment, with a total of only three (pre) or four (post) leading to execution with a research question or similar. With regard to the moderation of the execution of the experiments, the results show that the time in which pupils were given observation orders and got observation time more or less doubled, while all the other parameters approximately halved. Not only did the activating time increase, but the prospective teachers also paid much more attention to the respective sequences, so that the observation period corresponded much better to the action in the experiment. The time share spent explaining also decreased, although not as much as hoped. The prospective teachers also seemed to have difficulties to refrain from explanations during the presentation. But with the more intensive study of one's own linguistic guidance the prospective teachers' moderations became steadily shorter and more concise in content. This was reflected in the decrease of complaints that the playtime of the videos was too short. Nevertheless, some moderations remained long-winded. However, the linguistic content analysis of the moderations is still pending.

Training with feedback through eye tracking and assessment resulted in a significant increase in pupil activating time (RQ2). Our results of the interviews and surveys show that training with eye tracking as a feedback tool has a high level of acceptance and perceived usefulness among the prospective teachers. In the interviews, the prospective teachers described, among other things, how eye tracking feedback made them aware of their previous abilities to accompany experiments linguistically. The direct feedback from the pupils set in motion a process that made them realize the value of a good description of the experimental set-up but also the possibilities of attention-grabbing work assignments. This feedback acted back on our prospective teachers as described in the Process Model of SFT (Röhl et al., 2021). Furthermore, the eye tracking feedback with the accompanying verbal analysis by the lecturer provided prospective teachers with information on how to improve their moderation skills. The interviews revealed the extent to which prospective teachers grappled with this information and developed individual instructional approaches (see section "1.2.2. Levels and forms of feedback").

In the gaze overlay videos, when comparing pre- and post-moderation, one can clearly see the stronger focus of the pupils' gaze and the longer stay in one area. Unfortunately, this effect cannot be statistically represented in our approach since we only had three pupils per prospective teacher available. Watching the gaze overlay videos of their own moderation showed the prospective teachers the gap between the target (directing pupils' attention) and the current state.

6. Limitations

The results of the study should be interpreted with the following limitations. Firstly, the sample size is relatively small. Fifteen 5th semester prospective teachers participated in the study. At that time, they were all prospective teachers enrolled in that semester, at that location and on that course. Expanding the sample to include other locations or prospective teachers from other semesters might have led to biases in their prior knowledge or experience. In addition, the recorded video experiments of the pre-test and the post-test of all participating subjects were shown to three pupils each and their eye movements were recorded. The effort involved was already very high and would have increased massively with a larger sample. This was therefore not done. Secondly, the pupils' eye movements show the participants how their attentional cues work on the pupils' visual attention. Of course, they do not provide any information about what the pupils have actually learned. The changes in the pupils' knowledge were not the purpose of this study, but the reactions of them to the verbal input of prospective teachers providing feedback.

7. Conclusion

Observing the gaze behavior of pupils watching a "silent video "moderated by prospective teachers themselves gives them authentic feedback on their own effectiveness. Prospective teachers can literally see the impact of their words, the reaction of the pupils listening to them. They see where the pupils are looking on and individually recognize when or why the pupils leave the currently important areas of the set-up. The most important achievement, however, is that all prospective teachers can directly see and experience their own individual learning progress to accompany experiments in an attention-activating verbal way. They can see how even small changes (e.g., giving a second surface feature or describing the function) in moderating an experiment can have a lasting impact on pupils' attention.

The analysis of the connection between control codes (given as verbal attentional cues) and pupils' response leads to another important result of the use of eye tracking: pupils follow the command to look at a particular area of the set-up immediately almost every time, but as quickly as they look, they leave it again. To keep their attention on the spot, it is necessary to give a second assignment or to describe another feature such as the function (given as verbal content-related cues) or another surface feature of an object. Pupils who have received at least two pieces of information or assignments stay longer on this area of the set-up. To stay longer on a certain spot is very necessary for the pupils to make the observation the teacher intended. Thus, the results show that one strength of verbal cues, namely, being able to offer attentional guidance and content support, should also be used. With the more intensive study of one's own linguistic guidance the prospective teachers' moderations became steadily shorter and more concise in content. However, the linguistic content analysis of the moderations is still pending.

Demonstrating experiments in class in a way that is effective for learning requires a lot of practice. Training with "silent videos" is a promising method to support this exercise process, although it cannot replace real-life execution. It is not the intent of this training to standardize prospective teacher moderation, just as there is no ideal type of moderation, but rather everyone should develop their own individual appropriate teaching style. However, using eye tracking feedback gives prospective teachers unbiased and direct feedback from real pupils on their verbal skills and the impact of their use of language on their pupils. This study has so far analyzed only the group that received both the training and the eye tracking feedback. It is therefore unclear how much of the prospective teachers' positive development can be attributed to the training or the eye tracking feedback. Further research should explore how much of the improvement in the video presentations can be explained by the eye tracking feedback factor and how much by the training factor.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MS, BW, and RG contributed to the conception and design of the study. MS organized the database and wrote the first draft of the manuscript. MS and BW performed the statistical analysis. BW wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work "silent videos" of the Chair for Physics Education at LMU Munich is part of the project Lehrerbildung@LMU run by the Munich Center for Teacher Education, which was funded by the Federal Ministry of Education and Research under the funding code 01JA1810.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2023.1140272/full#supplementary-material

References

Addleman, D. A., and Jiang, Y. V. (2019). Experience-driven auditory attention. *Trends Cogn. Sci.* 23, 927–937. doi: 10.1016/j.tics.2019.08.002

Alemdag, E., and Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Comput. Educ.* 125, 413–428. doi: 10.1016/j.compedu.2018.06.023

Alpizar, D., Adesope, O. O., and Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educ. Tech. Res. Dev.* 68, 2095–2119. doi: 10.1007/s11423-020-09748-7

Amso, D. (2016). *Visual attention: Its role in memory and development*. Washington, DC: American Psychological Association. Available online at: https://dcnlab.psychology.columbia.edu/sites/default/files/content/Visual-attention-Its-role-in-memory-and-development.pdf

Boucheix, J. M., and Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learn. Instr.* 20, 123–135. doi: 10.1016/j.learninstruc.2009.02.015

Cannon, M. D., and Witherspoon, R. (2005). Actionable feedback: Unlocking the power of learning and performance improvement. *Acad. Manage. Perspect.* 19, 120–134. doi: 10.5465/ame.2005.16965107

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6:284. doi: 10.1037/1040-3590.6.4.284

Cullipher, S., Hansen, S. J., and VandenPlas, J. R. (2018). "Eye tracking as a research tool: An introduction," in *Eye tracking for the chemistry education researcher*, eds J. R. VandenPlas, S. J. R. Hansen, and S. Cullipher (Washington, DC: American Chemical Society), 1–9. doi: 10.1021/bk-2018-1292.ch001

de Koning, B. B., Tabbers, H. K., Rikers, R. M., and Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding?. *Learn. Instr.* 20, 111–122. doi: 10.1016/j.learninstruc.2009.02.010

Folyi, T., Fehér, B., and Horváth, J. (2012). Stimulus-focused attention speeds up auditory processing. *Int. J. Psychophysiol.* 84, 155–163. doi: 10.1016/j.ijpsycho.2012.02.001

Ginns, P. (2005). Meta-analysis of the modality effect. *Learn. Instr.* 15, 313–331. doi: 10.1016/j.learninstruc.2005.07.001

Glaser, M., and Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *J. Educ. Psychol.* 107:1006. doi: 10.1037/edu0000044

Hansen, S. J. R., Hu, B., Riedlova, D., Kelly, R. M., Akaygun, S., and Villalta-Cerdas, A. (2019). Critical consumption of chemistry visuals: Eye tracking structured variation and visual feedback of redox and precipitation reactions. *Chem. Educ. Res. Pract.* 20, 837–850. doi: 10.1039/C9RP00015A

Hattie, J. (2021). "Foreword," in *Student feedback on teaching in schools*, eds W. Rollett, H. Bijlsma, and S. Röhl (Cham: Springer), 1–7.

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Hu, J., and Zhang, J. (2021). The effect of cue labeling in multimedia learning: Evidence from eye tracking. *Front. Psychol.* 12:736922. doi: 10.3389/fpsyg.2021.736922

Jarodzka, H., Janssen, N., Kirschner, P. A., and Erkens, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative?. *Br. J. Educ. Technol.* 46, 803–817. doi: 10.1111/bjet.12174

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., and Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learn. Instr.* 25, 62–70. doi: 10.1016/j.learninstruc.2012.11.004

Katsuki, F., and Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *Neuroscientist* 20, 509–521. doi: 10.1177/1073858413514136

Keller, L., Cortina, K. S., Müller, K., and Miller, K. F. (2022). Noticing and weighing alternatives in the reflection of regular classroom teaching: Evidence of expertise using mobile eye tracking. *Instr. Sci.* 50, 251–272. doi: 10.1007/s11251-021-09570-5

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119:254. doi: 10.1037/0033-2909.119.2.254

Kriz, S., and Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *Int. J. Hum. Comput. Stud.* 65, 911–930.

Langner, A., Graulich, N., and Nied, M. (2022). Eye Tracking as a Promising Tool in Pre-Service Teacher Education- A New Approach to Promote Skills for Digital Multimedia Design. *J. Chem. Educ.* 99, 1651–1659. doi: 10.1021/acs.jchemed.1c01122

Lingzhu, J. (2003). Listening activities for effective top-down processing. *Internet TESL J.* 9.

Lockhofen, D. E. L., and Mulert, C. (2021). Neurochemistry of Visual Attention. *Front. Neurosci.* 15:643597. doi: 10.3389/fnins.2021.643597

Lupyan, G. (2017). Changing what you see by changing what you know: The role of attention. *Front. Psychol.* 8:553. doi: 10.3389/fpsyg.2017.00553

Mayer, R. E. (2014). "Cognitive theory of multimedia learning," in *The Cambridge handbook of multimedia learning*, ed. R. E. Mayer (Cambridge: Cambridge University Press), 43–71. doi: 10.1017/CBO9781139547369.005

Mayring, P. (2015). "Qualitative content analysis: theoretical background and procedures," in *Approaches to qualitative research in mathematics education. Advances in mathematics education*, eds A. Bikner-Ahsbahs, C. Knipping, and N. Presmeg (Dordrecht: Springer). Available online at: https://doi.org/10.1007/978-94-017-9181-6_13

McIntyre, N. A., and Foulsham, T. (2018). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instr. Sci.* 46, 435–455. doi: 10.1007/s11251-017-9445-x

McIntyre, N. A., Mainhard, M. T., and Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learn. Instr.* 49, 41–53. doi: 10.1016/j.learninstruc.2016.12.005

Minarikova, E., Smidekova, Z., Janik, M., and Holmqvist, K. (2021). Teachers' professional vision: Teachers' gaze during the act of teaching and after the event. *Front. Educ.* 6:716579. doi: 10.3389/feduc.2021.716579

Mussgnug, M., Lohmeyer, Q., and Meboldt, M. (2014). "Raising designers' awareness of user experience by mobile eye tracking records," in *DS 78: Proceedings of the 16th International conference on engineering and product design education (E&PDE14), design education and human technology relations*, (The Netherlands: University of Twente).

Narciss, S., and Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learn. Instr.* 16, 310–322. doi: 10.1016/j.learninstruc.2006.07.003

Ozcelik, E., Arslan-Ari, I., and Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Comput. Hum. Behav.* 26, 110–117. doi: 10.1016/j.chb.2009.09.001

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., and Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *J. Vis.* 13:16. doi: 10.1167/13.3.16

Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–13. doi: 10.1016/j.edurev.2015.12.003

Röhl, S., Bijlsma, H., and Rollett, W. (2021). "The process model of student feedback on teaching (SFT): A theoretical framework and introductory remarks," in *Student feedback on teaching in schools*, eds W. Rollett, H. Bijlsma, and S. Röhl (Cham: Springer). doi: 10.1007/978-3-030-75150-0_1

Rollett, W., Bijlsma, H., and Röhl, S. (2021). "Student feedback on teaching in schools: Current state of research and future perspectives," in *Student feedback on teaching in schools*, eds W. Rollett, H. Bijlsma, and S. Röhl (Cham: Springer), 259–270. doi: 10.1007/978-3-030-75150-0_16

Salvucci, D. D., and Goldberg, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on eye tracking research & applications*, (Palm Beach Gardens, FL: ACM), 71–78. doi: 10.1145/355017.355028

Schmidt-Weigand, F., Kohnert, A., and Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learn. Instr.* 20, 100–110. doi: 10.1016/j.learninstruc.2009.02.011

Schneider, S., Beege, M., Nebel, S., and Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educ. Res. Rev.* 23, 1–24. doi: 10.1016/j.edurev.2017.11.001

Schweinberger, M., and Girwidz, R. (2022). "Silent videoclips' for teacher enhancement and physics in class—material and training wheels," in *Physics teacher education*, eds J. Borg Marks, P. Galea, S. Gatt, and D. Sands (Cham: Springer), 149–159. doi: 10.1007/978-3-031-06193-6_11

Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795

Stuermer, K., Seidel, T., Mueller, K., Häusler, J., and Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft* 20, 75–92. doi: 10.1007/s11618-017-0731-9

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. doi: 10.1007/978-1-4419-8126-4

Szulewski, A., Braund, H., Egan, R., Gegenfurtner, A., Hall, A. K., Howes, D., et al. (2019). Starting to think like an expert: An analysis of resident cognitive processes during simulation-based resuscitation examinations. *Ann. Emerg. Med.* 74, 647–659. doi: 10.1016/j.annemergmed.2019.04.002

Watzka, B., Hoyer, C., Ertl, B., and Girwidz, R. (2021). Wirkung visueller und auditiver Hinweise auf die visuelle Aufmerksamkeit und Lernergebnisse beim Einsatz physikalischer Lern-videos. *Unterrichtswissenschaft* 49, 627–652. doi: 10.1007/s42010-021-00118-7

Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Front. Psychol.* 10:3087. doi: 10.3389/fpsyg.2019.03087

Xenos, M., and Rigou, M. (2019). Teaching HCI design in a flipped learning m. sc. course using eye tracking peer evaluation data. *arXiv* [preprint].

Xie, H., Mayer, R. E., Wang, F., and Zhou, Z. (2019). Coordinating visual and auditory cueing in multimedia learning. *J. Educ. Psychol.* 111:235. doi: 10.1037/edu0000285

TYPE Original Research
PUBLISHED 22 June 2023
DOI 10.3389/feduc.2023.1162281

OPEN ACCESS

EDITED BY
André Bresges,
University of Cologne, Germany

REVIEWED BY
Cédric Sueur,
UMR7178 Institut pluridisciplinaire Hubert Curien (IPHC), France
Bianca Watzka,
Otto von Guericke University Magdeburg, Germany

*CORRESPONDENCE
Larissa Hahn
✉ larissa.hahn@uni-goettingen.de

RECEIVED 09 February 2023
ACCEPTED 15 May 2023
PUBLISHED 22 June 2023

CITATION
Hahn L and Klein P (2023) Analysis of eye movements to study drawing in the context of vector fields. *Front. Educ.* 8:1162281.
doi: 10.3389/feduc.2023.1162281

Analysis of eye movements to study drawing in the context of vector fields

Larissa Hahn* and Pascal Klein

Physics Education Research, Faculty of Physics, University of Göttingen, Göttingen, Germany

Research has shown that visual representations can substantially enhance the learning and understanding of STEM concepts; despite this, students tend to struggle in using them fluently and consistently. Consequently, educators advocate for explicit instructions that support the coordination of multiple representations, especially when concepts become more abstract and complex. For recent years, the drawing (or sketching) technique has received increasing attention. Theoretical considerations and prior research suggest that drawing has the potential to support knowledge construction and to provide cognitive relief. In this article, we present two studies that investigate the impact of drawing activities in a multi-representational, instruction-based learning scenario from physics, more precisely, in the context of vector fields. Further, mobile and remote eye tracking was used to record students' gaze behavior in addition to monitoring indicators of performance and cognitive load. Here, eye movements provide information about cognitive processes during the completion of the instruction, on the one hand, and during subsequent problem solving, on the other hand. Comparisons of a treatment group instructed with drawing activities and a control group instructed without drawing activities revealed significant differences in students' perceived cognitive load ($p = 0.02$, $d = 0.47$ and $p = 0.0045$, $d = 0.37$), as well as their response accuracy ($p = 0.02$, $d = 0.51$) and their response confidence ($p = 0.02$, $d = 0.55$ and $p = 0.004$, $d = 0.64$) during assessment after instruction ($N = 84$). Moreover, students instructed with drawing activities were found to distribute more visual attention to important parts of the instruction (vector field diagram and instructional text, $N = 32$) compared to the control group and, further, showed effective, expert-like behaviors during subsequent problem solving ($N = 53$). Finally, as a contribution to current trends in eye-tracking research, the application of mobile and remote eye-tracking in drawing-based learning and assessment scenarios is compared and critically discussed.

KEYWORDS

drawing, multiple representations, physics, vector fields, eye movements, instruction-based learning, conceptual understanding, eye tracking

1. Introduction

Multiple external representations (MERs) and their interpretations play a central role in STEM and, particularly, in STEM learning. As different representations focus on different characteristics of a concept or a learning subject, and thus complement and constrain each other, multiple, visual representations enable a deep understanding of fundamental scientific concepts (Ainsworth, 1999; Seufert, 2003; Rau, 2017). Moreover, a flexible and conscious use of different representations, which requires enhanced representational competencies, was found to have positive effects on knowledge acquisition, development of conceptual understanding, and problem-solving skills (e.g., Nieminen et al., 2012; Chiu and Linn, 2014; Rau, 2017).

However, learning from and with multiple representations in STEM education often places special demands on the visuo-spatial working memory, hence increasing cognitive load (Baddeley, 1986; Cook, 2006; Logie, 2014). Therefore, current research advocates assistance through drawing activities, for example, by sketching (or drawing) visual cues or by transforming textual information into a drawing (e.g., Figure 1, Left; Kohnle et al., 2020; Ainsworth and Scheiter, 2021). As visualizations are integral to scientific thinking, Ainsworth et al. (2011) emphasized the potential of drawing as an effective learning strategy. Following the authors, the externalizing approach of drawing aligns with the visual-spatial demands of science learning, thus helping students to visually make sense of concepts and providing cognitive relief (Bilda and Gero, 2005; Wu and Rau, 2018). Previous studies have confirmed these theoretical considerations by demonstrating positive learning outcomes of sketching activities in multi-representational learning environments (Leopold and Leutner, 2012; Wu and Rau, 2018; Hellenbrand et al., 2019; Kohnle et al., 2020; Ainsworth and Scheiter, 2021). As such, sketching increased attention and engagement with the representation, allowed to pay more attention to details and important parts of a representation, and thus supported a (visual) understanding of concepts.

Studies investigating cognitive load often postulate limited capacity of working memory resources. This is typically broken down into three types of cognitive load—that are, intrinsic, extraneous, and germane cognitive load (Cognitive Load Theory; Sweller, 2010). Intrinsic cognitive load refers to the inherent complexity of the information to be understood (further specified below for the context of vector fields), extraneous cognitive load to the design of the instructional material, and germane cognitive load to the mental capacities a learner devotes to the learning subject. In that context, sketching activities are considered to allow a more effective use of visuo-spatial working memory resources (Bilda and Gero, 2005; Sweller, 2010). Moreover, studies have shown that using multiple external representations in STEM education, particularly with a focus on drawing, can enhance knowledge acquisition, conceptual understanding, and problem-solving skills while reducing cognitive load (Leopold and Leutner, 2012; Wu and Rau, 2018; Hellenbrand et al., 2019).

In recent years, eye-tracking as a nonintrusive process-based assessment technique has become increasingly popular in research on (multimedia) learning (Hyönä, 2010; Lai et al., 2013; Alemdag and Cagiltay, 2018) and STEM education (e.g., reviews in physics and mathematics see Strohmaier et al., 2020; Hahn and Klein, 2022a). Following the eye-mind hypothesis by Just and Carpenter (1976), refined by Wu and Liu (2022), many studies postulate a relationship between visual attention and cognitive processing (e.g., Tsai et al., 2012). In this context, eye-tracking measures, such as total and mean fixation duration, fixation and transition count, and time to first fixation, are typically used to study students' attention and visual behavior during learning and assessment activities. This includes analyzing their engagement with instructional designs, their strategies during problem solving, and performance on concept tests or tasks (Lai et al., 2013; Hahn and Klein, 2022a), thus allowing to study expertise differences (Gegenfurtner et al., 2011). Thereby, research focused on cognitive processes in learning or assessment scenarios; studies analyzing gaze data during instruction and subsequent assessment, in contrast, were hardly conducted (Hahn and Klein, 2022a). In physics education research, gaze data is typically collected on a screen using remote eye-tracking methods; only one study used mobile eye-tracking glasses thus far (Hahn and Klein, 2022a). As, particularly paper-based, drawing activities significantly influence learning behaviors, mobile eye tracking promises to be particularly suited to track its constructive mechanisms and to gain insight into frequency and timing of externalization (Hellenbrand et al., 2019).

In university physics, vector field representations play a central role, being represented either algebraically as a formula or graphically as a vector field diagram (see Figure 1). When representing a quantity as a vector field, the field's divergence, a measure for its sources and sinks, is of particular relevance for physics applications (Griffiths, 2013). For electrodynamics, one of the largest subfields of physics, an extensive preparation in vector calculus was found to be highly correlated with students' performance at university (Burkholder et al., 2021). However, further research also revealed that students often lack a conceptual understanding of vector field representations and, particularly, divergence, which is highly relevant to physics comprehension (e.g., Pepper et al., 2012; Singh and Maries, 2013; Bollen et al., 2015). Recent studies found, for example, that students interpreted the divergence of a vector field literally instead of referring to its physics-mathematical concepts and that they struggled with evaluating the divergence from a vector field diagram (Ambrose, 2004; Pepper et al., 2012; Singh and Maries, 2013; Bollen et al., 2015, 2016, 2018; Baily et al., 2016; Klein et al., 2018, 2019, 2021). Particularly surprising, graduate students struggle with conceptual aspects of divergence, even though they know how to calculate it mathematically (Singh and Maries, 2013). Several studies deepened this line of research and identified various learning difficulties closely related to the Cartesian representation of divergence. In particular, the concept of covariation between field components and coordinates was found to cause difficulties for students (see Figure 1, Left); they confused components with coordinates or committed errors when decomposing vectors into their components (e.g., Gire and Price, 2012; Barniol and Zavala, 2014; Bollen et al., 2017). Additionally, Pepper et al. (2012) reported about student problems in dealing with partial vector derivatives as they, for example, confused the change of a vector with its magnitude. Analysis of students' gaze when viewing vector field diagrams confirmed the difficulties mentioned above (Klein et al., 2018, 2019, 2021). Moreover, it was shown that conceptual gaps regarding vector calculus were transferred to errors in application (e.g., in electrostatics and -magnetism; Ambrose, 2004; Jung and Lee, 2012; Bollen et al., 2015, 2016; Li and Singh, 2017).

Analysis of introductory and advanced physics texts and textbooks by Smith (2014) revealed that divergence is typically introduced using a mathematical expression, but is either not or insufficiently explained or discussed qualitatively or illustratively. In light of the aforementioned empirical findings, several researchers advocated new instructions using visual representations that address a conceptual understanding of vector field concepts and, particularly, divergence. Following this line of research, Klein et al. (2018) developed text-based instructions for visually interpreting divergence using vector field diagrams. Here, the authors referred to two different mathematical approaches, namely flux through boundary (integral approach)

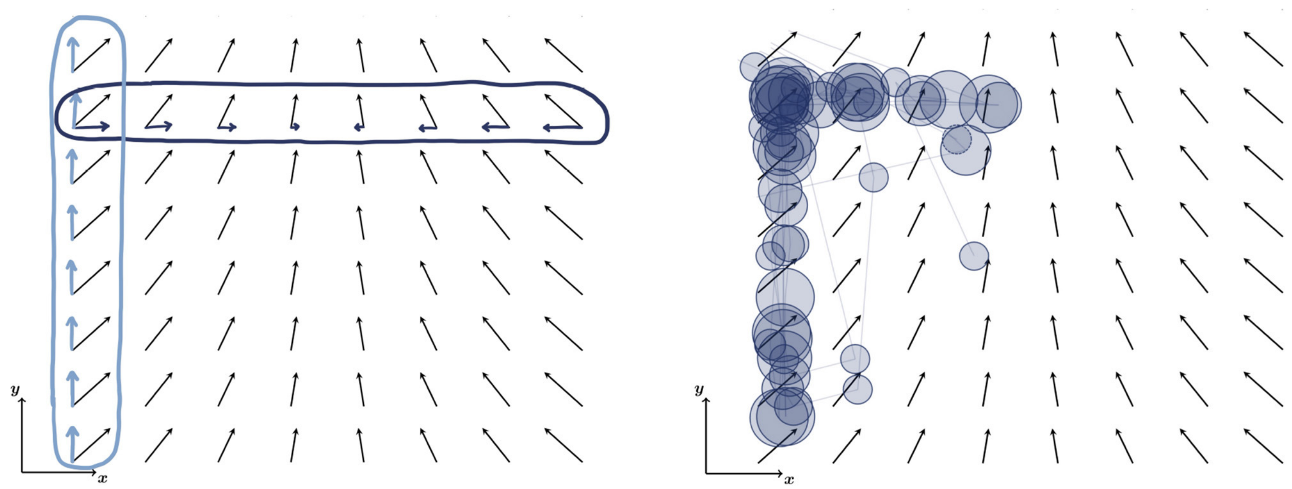


FIGURE 1
Vector field diagram of the field $\vec{F}(x, y) = -x\,\hat{e}_x + \hat{e}_y$ with drawing of visual cues and vector field components **(Left)** and visual evaluation in horizontal and vertical directions **(Right)** to interpret covariation in Cartesian coordinates. x and y components of the field are sketched in dark and light blue, respectively **(Left)**. To evaluate their change along a row (x direction) and a column (y direction), the corresponding row or column is circled. Analysis of covariation yields that the x component of the field changes in x direction (along the row), while the y component is constant in y direction (along the column). Visual evaluation of covariation along Cartesian directions (x and y direction) is associated with horizontal and vertical saccades (Scanpath; fixations are visualized by circles; circle size relates to fixation duration; **Right**).

and covariation of components and coordinates (differential approach). A clinical eye-tracking study revealed a quantitative increase in conceptual understanding after students completed the instructions (Klein et al., 2018). However, in post-intervention interviews, some subjects indicated difficulty with diagram-specific mental operations, such as decomposing vectors and evaluating field components along coordinate directions. They further suggested visual aids, for example, sketches of component decomposition, as being helpful to improve their performance. Hence, in a follow-up experimental study, Klein et al. (2019) compared two instructions of the differential strategy, with and without visual cues, and found that adding visual cues for component decomposition actually led to better learning outcomes. Moreover, a positive correlation with students' response confidence was found, as students from the treatment group instructed with visual cues trusted their answers more (Lindsey and Nagel, 2015; Klein et al., 2017, 2019). However, students' transfer performance and their perceived task difficulty during problem-solving did not improve significantly compared to the previous study indicating that visual cues did not fully overcome students' difficulties regarding vector decomposition and partial derivatives. In a third study, using a very similar instruction, Klein et al. (2021) found that particularly students with high or medium spatial abilities—as measured by the Spatial Span Task (SST) by Shah and Miyake (1996)— benefited from the instructional support, whereas students with low spatial abilities perceived high cognitive load and hardly profited from the instruction.

In addition to performance measures, all three studies used eye tracking to analyze students' handling of the instruction and representation-specific visual behaviors during subsequent problem-solving (Klein et al., 2018, 2019, 2021). In the studies by Klein et al. (2019, 2021), it was shown that visual cues increased fixation count and total fixation duration on relevant parts of the instruction. Furthermore, cognitive integration processes indicated by transitions were significantly more pronounced for students instructed with cues. During problem solving, saccadic eye movement patterns of students instructed with visual cues were similar to experts. Here, the authors referred to results by Klein et al. (2018), who found that best-performing students' eye movements are dominated by horizontal and vertical saccades indicating correct interpretation of partial derivatives along Cartesian coordinate directions (Figure 1, Right). Besides characteristic saccadic directions, Klein et al. (2018) found that best-performing students also performed shorter saccades which was associated with a systematic evaluation of component changes in the vector field diagram. However, despite high visuo-spatial demands of the task stimulus, no correlation of such gaze patterns with students' spatial abilities could be confirmed (Klein et al., 2021).

By taking the above-mentioned theoretical considerations into account, the aforementioned multi-representational instructions on divergence (Klein et al., 2018, 2019, 2021) were evolved and extended by dedicated pre-exercises on vector decomposition and partial derivatives. In particular, several task on drawing of vector components or highlighting rows and columns to support evaluation along coordinate directions are incorporated (Hahn and Klein, 2021). The tasks aim at providing cognitive relief, fostering further engagement with the representations and the instructed strategy, and supporting the development of a conceptual understanding of divergence which can be transferred to further concepts. Besides analyzing the general impact of the instruction, this contribution aims at evaluating the value of the drawing activities in a multi-representational instruction on divergence by presenting two studies both comparing a treatment group instructed with drawing activities and a control group instructed without drawing activities. The following guiding question is investigated: "Does the instruction on divergence including drawing activities have a higher learning impact than the instruction without drawing activities?" Besides typical performance indicators, such as response accuracy and response confidence, eye movements are exploited to reveal between-subjects effects during completion of the instruction and subsequent problem solving. Consequently, considering theoretical frameworks on the use of drawing activities, the following research questions and hypotheses are posed:

(RQ1) What impact do drawing activities have on students' conceptual understanding and their accuracy of judging the divergence of vector field diagrams?
(RQ2) What are the differences between students instructed with and without drawing activities concerning their performance (response accuracy and confidence) as measured by several assessment tasks (including the evaluation of the divergence of vector field diagrams, the interpretation of partial derivatives of vector field diagrams, and conceptual questions on divergence)?
(RQ3) What are the differences between students instructed with and without drawing activities concerning their cognitive load (as measured by a cognitive load questionnaire) during learning and task processing?
(H1) Students instructed with drawing activities perceive less cognitive load (as measured by a cognitive load questionnaire) during learning and task processing.
(RQ4) What are the differences between students instructed with and without drawing activities concerning their visual attention
(a) during processing of the instruction?
(b) when evaluating the divergence of vector field diagrams?

2. Materials and methods

The aforementioned research questions are investigated in two separate studies. Study 1 uses performance measures to quantify the impact of the instruction and, particularly, the drawing activities in order to address research questions RQ1, RQ2 and RQ3 including hypothesis H1. In study 2 cognitive processing is focused using mobile and remote eye tracking during completion of the instruction and subsequent problem solving, thus addressing research questions RQ4(a) and RQ4(b).

2.1. Study 1

2.1.1. Participants

The sample of study 1 was drawn from physics students at a German university (Table 1) in the context of a large-scale voluntary physics pre-course before the first study semester. Prior to the study, students received a short introduction to vector fields in the lecture. In the corresponding recitations, students completed the study material in self study (for study design see Section 2.1.2 and Figure 2, and for materials see

TABLE 1 Sample data of study 1 (Left) and study 2 (Right; treatment group TG, control group CG, number No.).

	Study 1			Study 2		
	Total	TG	CG	Total	TG	CG
Number of subjects	84	43	41	54	27	27
No. of female subjects	19	5	14	17	9	8
Age range (in years)	18 – 23	18 – 21	18 – 23	18 – 26	18 – 26	18 – 22
Mean age (in years)	19.2 ± 0.9	19.2 ± 0.7	19.2 ± 1.0	20.2 ± 1.9	20.2 ± 1.9	20.2 ± 1.8
No. of semesters studied	1	1	1	2.7 ± 1.3	2.6 ± 1.1	2.9 ± 1.5
Average grade for university entrance qualification[a]	1.6 ± 0.5	1.6 ± 0.5	1.6 ± 0.6	1.7 ± 0.7	1.7 ± 0.6	1.7 ± 0.8
No. of subjects with advanced maths course at school	72 (86%)	33 (77%)	39 (95%)	35 (65%)	15 (56%)	20 (75%)
No. of subjects with physics course at school	69 (82%)	38 (88%)	31 (76%)	41 (76%)	20 (74%)	21 (78%)
School performance in mathematics[b]	8.0 ± 1.5	7.8 ± 1.6	8.1 ± 1.4	8.3 ± 1.3	8.4 ± 1.4	8.2 ± 1.3
School performance in physics[b]	8.3 ± 1.3	8.3 ± 1.4	8.4 ± 1.2	8.0 ± 1.6	8.0 ± 1.6	8.1 ± 1.7
Spatial abilities[c]	–	–	–	0.54 ± 0.18	0.55 ± 0.16	0.51 ± 0.22
Pretest score vector fields[d]	0.93 ± 0.17	0.94 ± 0.17	0.93 ± 0.17	0.97 ± 0.13	0.94 ± 0.17	0.99 ± 0.06
Pretest score divergence[d]	0.76 ± 0.23	0.74 ± 0.25	0.79 ± 0.21	0.73 ± 0.24	0.69 ± 0.25	0.77 ± 0.23
No. of correct answers for vector field VF_0	33 (39%)	16 (37%)	17 (41%)	22 (41%)	12 (44%)	10 (37%)

[a]The scale ranges from 1.0 (best performance) to 4.0. The grades are indicated by the students.
[b]The scale ranges from 0 to 10 (best performance). The scores are based on students' self-assessment.
[c]The scale ranges from 0 to 1 (best performance), measured by the spatial span task (conventional score).
[d]The scale ranges from 0 to 1 (best performance).

Section 2.3, Table 2, Figures 3, 4 and Supplementary material). In total, 84 first-year students (19 female, average age 19.2 years) participated in the study (for further characterization of the sample see Table 1, Left). It is notable that the pretest scores on vector fields are rather high, indicating sufficient prior knowledge about visual representations of vector fields and decomposition of single vectors into components of all students to understand the subsequent instruction. But as only 39% of them were able to evaluate the divergence of a vector field diagram the instruction could still have a meaningful impact (Table 1).

2.1.2. Study design, procedure, and data analysis methods

Study 1 uses a mixed design including within- and between-subjects treatments (Figure 2). The study procedure is summarized in Table 2 including an overview of all instruments and data (see Section 2.3 and the Supplementary material). First, students took a prior knowledge test, where they were asked to judge whether field components in a vector field diagram equal zero or not, and a pretest that included conceptual questions about divergence. The students then completed a three-page instruction including a pre-exercise on vector decomposition and partial derivatives (Figure 3) and an instruction on divergence (Figure 4) either with drawing activities (treatment group TG) or without drawing activities (control group CG; instructional material adapted from Klein et al., 2018, 2019, 2021). The group assignment was randomized by recitation group, where students selected a fixed recitation group by their own without knowing about the assignment to a treatment condition and about other group members. After finishing the instruction, the students took the posttest. It included several problem-solving tasks with vector field plots (Figure 5), several transfer tasks on partial derivatives, and the conceptual pretest questions on divergence. Last, students' perceived cognitive load and sociodemographic data have been collected. No significant differences between the two groups (treatment and control group) regarding various sociodemographic data and performance indicators were found (Table 1, Left).

For data analysis, within-subjects effects are investigated through pre and post comparisons of students' achievement in the concept test on divergence and in judging a fields' divergence (initial problem VF_0 and VF_5 from the problem-solving phase). Here, standard methods of quantitative statistics (e.g., *t*-tests) were used by referring to the interpretation of Cohen (1988). Additionally, treatment and control group were statistically compared with respect to students' perceived cognitive load as well as their response accuracy and confidence during problem solving, in the transfer task, and in the posttest.

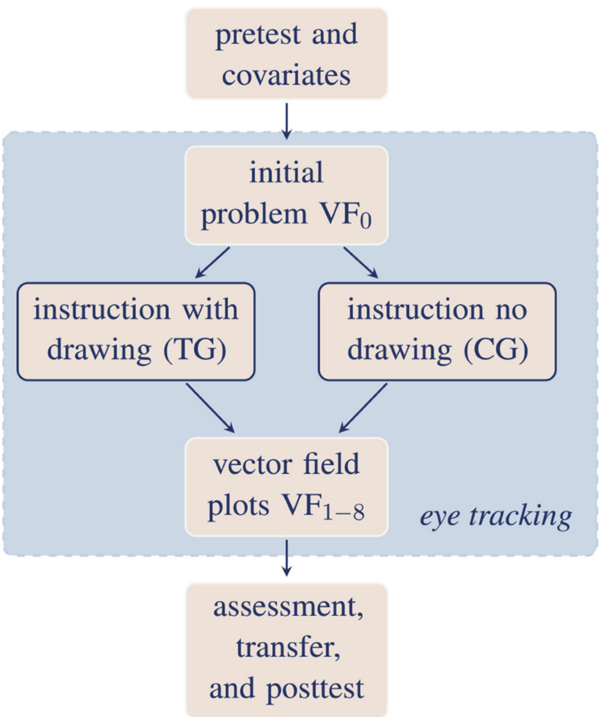


FIGURE 2
Study design for study 1 (beige) and 2 (extension in blue; treatment group TG, control group CG). All parts of study 1 are colored in beige. In study 2, for the initial problem VF_0, the instruction, and the problem-solving task after instruction additional eye-tracking data was collected (colored in blue).

2.2. Study 2

2.2.1. Participants

The sample of study 2 was drawn from physics students at the same German university than study 1 in the context of a large-scale second-semester physics lecture on electromagnetism. Prior to the study, students received a short introduction to vector fields in the context of electric and magnetic fields in the lecture. In total, 54 students (17 female, average age 20.2 years), mostly in their second year of study, participated in the study (for further characterization of the sample see Table 1, Right). Again, students' pretest scores indicated sufficient prior knowledge to understand the subsequent instruction.

2.2.2. Study design and procedure

Study 2 uses the same mixed design as study 1 including drawing activities as between-subjects treatment (Section 2.1.2, Figure 2, Table 2, and Figures 3, 4). Subjects participated voluntarily in the study and were compensated with 20€. Group assignment was randomized before the start of the study and students were guided individually to the eye-tracking laboratory. After prior knowledge assessment, a standardized test on spatial abilities was administered (Spatial Span Task; Shah and Miyake, 1996), which measures the ability to simultaneously process and hold spatial information in memory. In addition to the performance measures from study 1, eye tracking was used to capture gaze behavior during instruction and problem solving (Section 2.4). Again, no significant differences between treatment and control group regarding various sociodemographic data and performance indicators were found (Table 1, Right).

TABLE 2 Study procedure (including approximate time) and overview of instruments and data collected for study 1.

Phase	Description	Time (min)	Instruments and data collected (source and ptly. reliability of the instruments)
1	Pretest: Prior knowledge	10	Prior knowledge test on vector field components (6 items; Klein et al., 2018, 2019, 2021): • response accuracy ($\alpha = 0.76$) • response confidence ($\alpha = 0.98$) Concept test on divergence (6 items; Baily et al., 2016; Bollen et al., 2018; Klein et al., 2018, 2019, 2021; Hahn and Klein, 2022b): • response accuracy ($\alpha = 0.54$) • response confidence ($\alpha = 0.88$)
2	*Spatial abilities*	*5*	*Spatial Span Task (Shah and Miyake, 1996)*
3	Initial problem VF_0	4	• response accuracy • *gaze data (mobile)*
4	Instruction (see Figures 3, 4)	15	• response accuracy (pre-exercise; $\alpha = 0.48$) • *gaze data (mobile)*
5	Problem solving (see Figure 5)	7	Eight vector field plots (8 items; Klein et al., 2018, 2019, 2021): • response accuracy ($\alpha = 0.77$) • response confidence ($\alpha = 0.98$) • *gaze data (mobile and remote)*
6	Assessment	5	Mental demands of task and instruction (Leppink et al., 2013; Klein et al., 2018, 2019, 2021): • Intrinsic Cognitive Load (3 items, $\alpha = 0.88$) • Extrinsic Cognitive Load (3 items, $\alpha = 0.66$) • Germane Cognitive Load (3 items, $\alpha = 0.87$)
7	Transfer (see Figure 5)	8	Three vector field plots (12 items; Klein et al., 2018, 2019, 2021): • response accuracy ($\alpha = 0.86$) • response confidence ($\alpha = 0.99$)
8	Posttest	5	Concept test on divergence (6 items; Baily et al., 2016; Bollen et al., 2018; Klein et al., 2018, 2019, 2021; Hahn and Klein, 2022b): • response accuracy ($\alpha = 0.61$) • response confidence ($\alpha = 0.86$)
9	Sociodemographics	5	Questionnaire

For relevant data and variables, reliability of the instruments is also indicated (Cronbach's alpha α). Additional (eye-tracking) data collection in study 2 is indicated in italics. For detailed description of instruments and materials used in phases 1–8 see Sections 2.1.2 and 2.3 and Supplementary material.

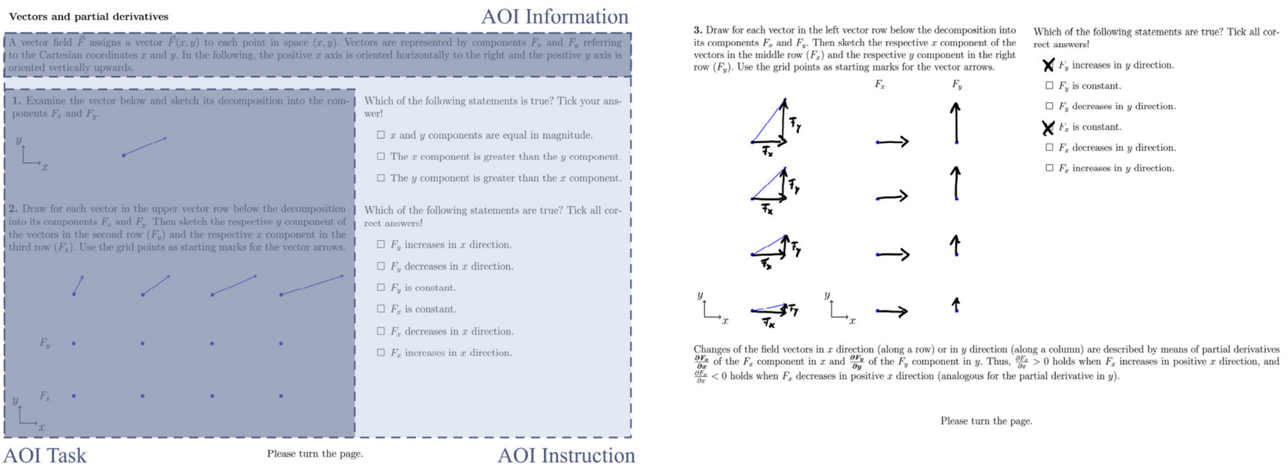


FIGURE 3
Two-sided pre-exercise on vector components and partial derivatives with drawing activities (treatment group) translated from originally German (the original instruction can be found in the Supplementary material). The second pre-exercise page shows an exemplary completion of the task **(Right)**. In the exercise without drawing activities, the decomposition tasks (left side of the page, respectively) are to be completed without drawing vectors (control group). Definition of AOIs for the first pre-exercise page is marked in blue **(Left)**. The AOIs for both pages are defined identically covering the whole instruction (AOI Instruction), the decomposition tasks (AOI Task), and a definition/information about vector fields or partial derivatives (AOI Information), respectively.

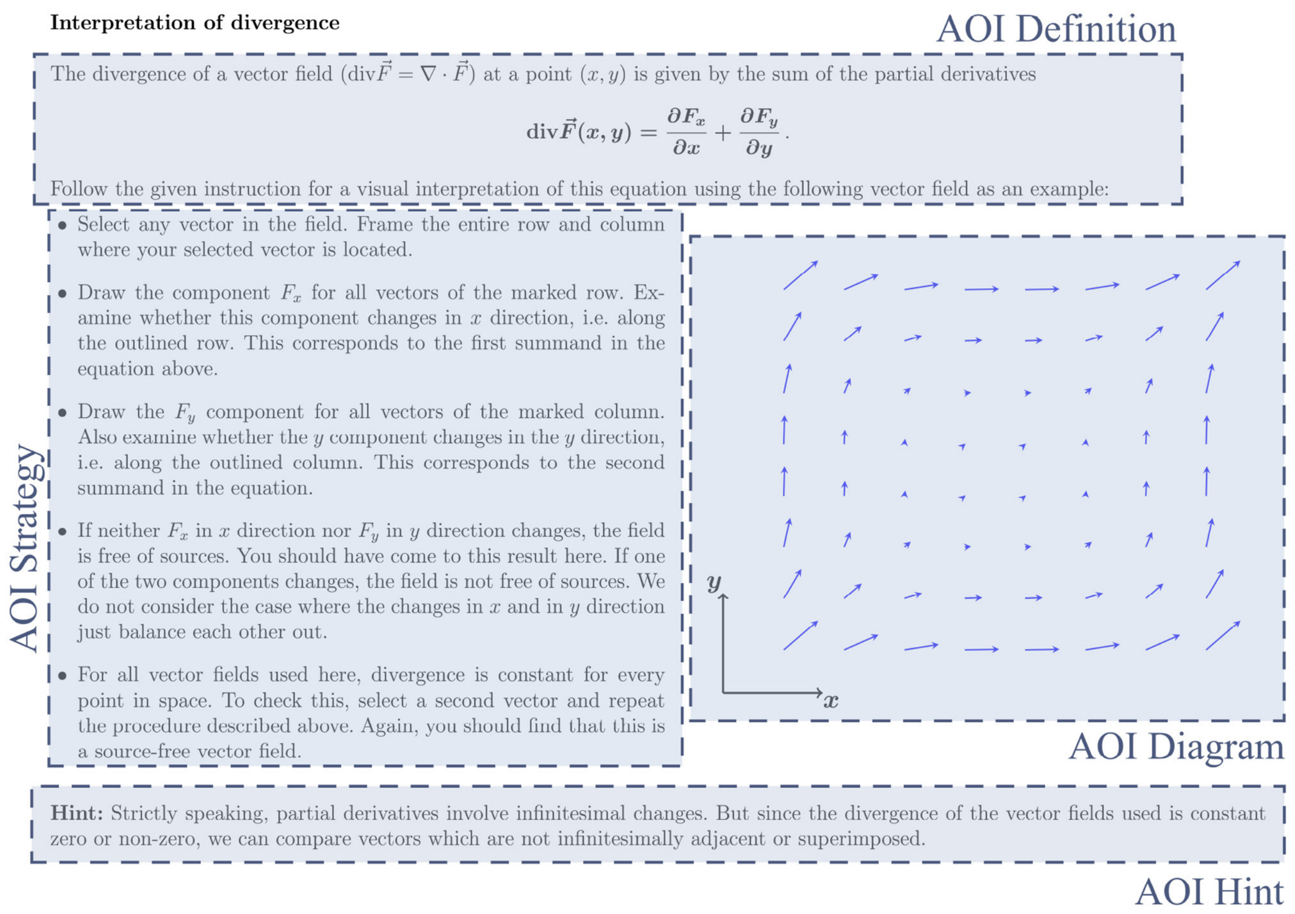


FIGURE 4
Instruction on visually interpreting divergence of vector fields with drawing activities (treatment group) translated from originally German (the original instruction can be found in the Supplementary material). The explanatory text aims at drawing vector components into the adjacent vector field diagram in order to infer the fields' divergence. In the instruction without drawing activities, the explanatory text asks for a mental decomposition of the vectors (control group). The defined AOIs for eye-tracking analysis are marked in blue. The AOIs follow the structure by Klein et al. (2019) covering the definition of divergence (AOI Definition), the instructional text (AOI Strategy), the vector field plot (AOI Diagram), and the concluding note (AOI Hint).

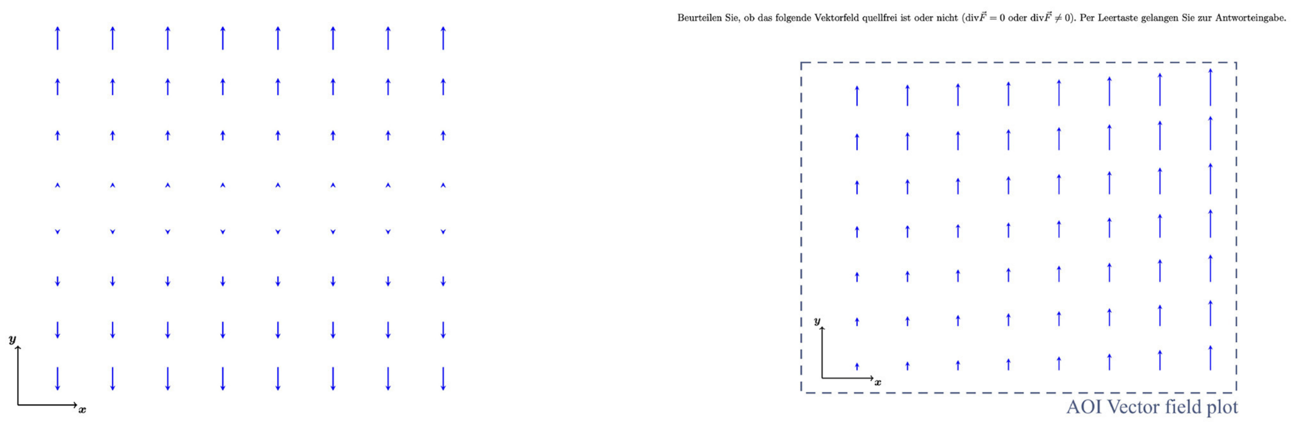


FIGURE 5
Examples of vector field diagrams used in the problem-solving tasks and the transfer tasks (**Left**; further development of Klein et al., 2018, 2019, 2021) and AOI definition **(Right)**.

2.3. Materials and measures

Pre-exercises and divergence instruction. The instruction on visually interpreting the divergence of vector fields included preliminary exercises addressing component decomposition and assessment of component change along an isolated row or column, in other words, partial derivatives (Figure 3). The following instruction on a visual interpretation of divergence consisted of a short introduction, an explanatory text with an adjacent vector field diagram, and a concluding note (Figure 4). The pre-exercise and the divergence instruction differed between treatment and control group in that students in the treatment group were asked to draw the vector components, while students in the control group constructed them mentally. Based on the sketched or mentally represented components, the changes of the components in the direction of the Cartesian coordinates (the partial derivatives), had to be evaluated (pre-exercises) and conclusions about the fields' divergence had to be drawn (divergence instruction). Construction and design of the instruction were based on materials used in prior studies (Klein et al., 2018, 2019, 2021).

Problem solving and transfer: Vector field plots. The vector field diagrams used in the study met certain requirements (Singh and Maries, 2013; Klein et al., 2018); first, the vector fields were created not to reflect any physical reality in order to exclude recognition effects. Second, the length scales were to be interpreted arbitrarily (any non-zero constant equals 1 or -1 by definition), and third, the dependencies of the vector field components were at most linear. Last, all vector fields were embedded in two-dimensional Cartesian coordinate systems, represented with approximately 25 arrows. Examples are given in Figure 5. For the problem-solving tasks, students were asked to indicate if a given vector field plot has zero or non-zero divergence. In the transfer task, it was asked for weather the partial derivatives are < 0, > 0, or $= 0$. In study 2, the first four vector field plots VF_{1-4} as well as the second four vector field plots VF_{5-8} were designed in parallel.

Divergence concept test. For assessment of conceptual knowledge regarding the divergence of vector fields, a concept test including six items was deployed. The items were designed in multiple-choice or true-false format. Most of them were taken from established concept tests on electrodynamics (CURrENT) or have been used and validated in a similar form in previous studies (Baily et al., 2016; Bollen et al., 2018; Klein et al., 2018, 2019, 2021). An example item is "The divergence can be different for every spot in the field" (Baily et al., 2016; translated for the study into German).

Questionnaire on cognitive load. The questionnaire addressing the mental demands of instruction and task processing consisted of nine items from an established instrument for measuring cognitive load (Leppink et al., 2013). All three types of cognitive load are addressed by three items each. A fourth item, which was originally dedicated to measure germane cognitive load, was omitted due to insufficient fit.

For both studies, the same materials were used.

2.4. Eye-tracking procedure, areas of interest, and data analysis methods (study 2)

Eye-tracking data collection was two-fold: Mobile eye tracking was used during instruction reading and solving the first four vector field problem-solving tasks VF_{1-4} and remote eye-tracking was used while working on the last four vector field diagrams VF_{5-8}. While remote eye-tracking was used to collect data for all 54 students, in the mobile eye-tracking phase due to technical capacities only the gaze of 33 students was tracked.

Mobile gaze data was recorded using wearable eye-tracking glasses from Tobii (Tobii Pro Glasses 3) with a sampling rate of 50 Hz and an accuracy of 0.6° visual angle. Tobii Glasses software and a one-point system-controlled calibration including a calibration card was used for data collection. Further, all materials were provided on large-scale paper sheets and on a tripod to enable a perpendicular viewing angle. One participant had to be discarded from analysis due to data loss caused by technical issues.

The last four problem-solving tasks were presented on a 24-inch computer screen (1920 × 1080 pixel resolution, 60 Hz frame rate) and eye movements were recorded using a stationary head-free eye-tracking system from Tobii (Tobii Pro Fusion). The eye tracker operates with an accuracy of less than 0.3° visual angle and a sampling frequency of 120 Hz. A 9-point calibration was used and agreement between the measured gaze positions and the actual points on the screen was checked by the experimenter. Calibration was repeated if the accuracy result was not satisfactory. A calibration plot showed error bars for each of the nine positions, indicating the differences between the gaze point calculated by the eye tracker and the actual dot position. When the eye tracker could not detect enough calibration data, the participant was re-positioned in front of the eye tracker and checked for any factors that could have been interfering with pupil detection. After careful checking of calibration results, one participant had to be discarded due to insufficient fit. Average distance between participant and screen was 64 cm.

For all gaze data, the distinction between fixations and saccades was made using a velocity threshold of $30°/s$ (Tobii I-VT Fixation Filter; Olsen, 2012). Data visualization and analysis was performed using Tobii Pro Lab 1.204 software. First, mobile eye-tracking data was mapped to the corresponding instruction and problem-solving pages by also marking whenever students were focusing outside the worksheet, on their hand, or the answer sheet. Then, areas of interest (AOIs) were defined for quantitative analysis of all three instruction pages (Figures 3, 4). For gaze-data analysis during problem solving, one area of interest covering the vector field diagram was defined for all eight tasks (Figure 5, Right). Due to the identical design of the problem-solving pages, same-sized AOIs could be placed at the same position for all eight tasks.

To investigate differences in cognitive processing of the instruction, fixation count, mean fixation duration, and time to first fixation are compared between both groups (with and without drawing activities) by analyzing visual attention distribution on the defined AOIs of the pre-exercises and the divergence instruction.

As total fixation duration and fixation count were found to be dependent, thus showing analogous effects, only fixation count is reported (Susac et al., 2019). Furthermore, to analyze gaze behavior on the vector field diagram during problem solving, process-based metrics—that are, proportion and length of horizontal and vertical saccades within a tolerance margin of $\pm 5°$—are analyzed. For mobile eye tracking, saccades (and transitions) were mostly not tracked correctly as the eye-tracking glasses seem to lose track during gaze movement. This was indicated by comments such as "EyesNotFoundMovement" and "UnknownEyeMovement" in the raw gaze data. Thus, for mobile gaze-data analysis, only fixation-based metrics—that are, fixation count, mean fixation duration, and time to first fixation—are used (see Discussion Section 4.3). All comparisons are conducted by use of standard methods of quantitative statistics (e.g., t-tests) and by referring to the interpretations of Cohen (1988).

3. Results

3.1. Study 1

Students' performance improved after instruction, as shown by the increase in their accuracy in the concept test (from 0.76 ± 0.23 to 0.85 ± 0.21) and in solving the initial problem (from 0.39 ± 0.49 to

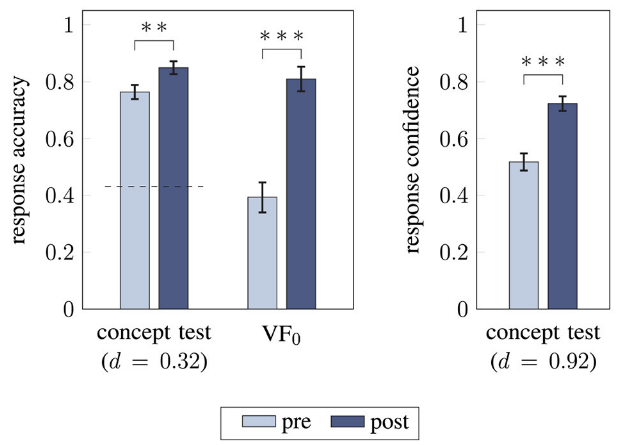


FIGURE 6
Comparison of students' performance before and after instruction using paired t-tests and McNemar-test (two-sided). Response accuracy for concept test and initial problem **(Left)** as well as response confidence for the concept test **(Right)** are compared between pre and post (**/*** statistical significance ($p < 0.01$ / $p < 0.001$), effect size d, the dashed line indicates the guessing probability, error bars represent 1 SEM).

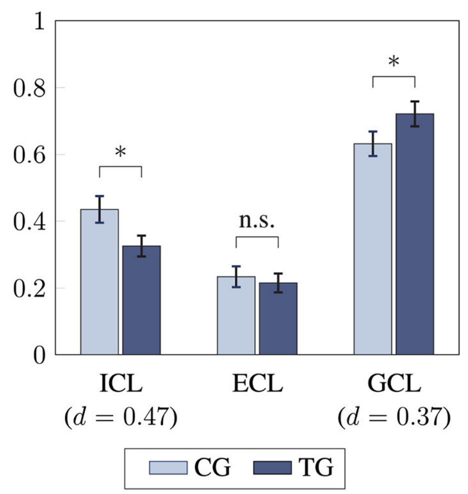


FIGURE 8
Comparison of students' perceived cognitive load in control and treatment group using unpaired t-tests (one-sided). Intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL) are compared between treatment group (TG) and control group (CG; * statistical significance ($p < 0.05$), n.s. "(statistically) not significant" indicating p-values > 0.05, effect size d, error bars represent 1 SEM).

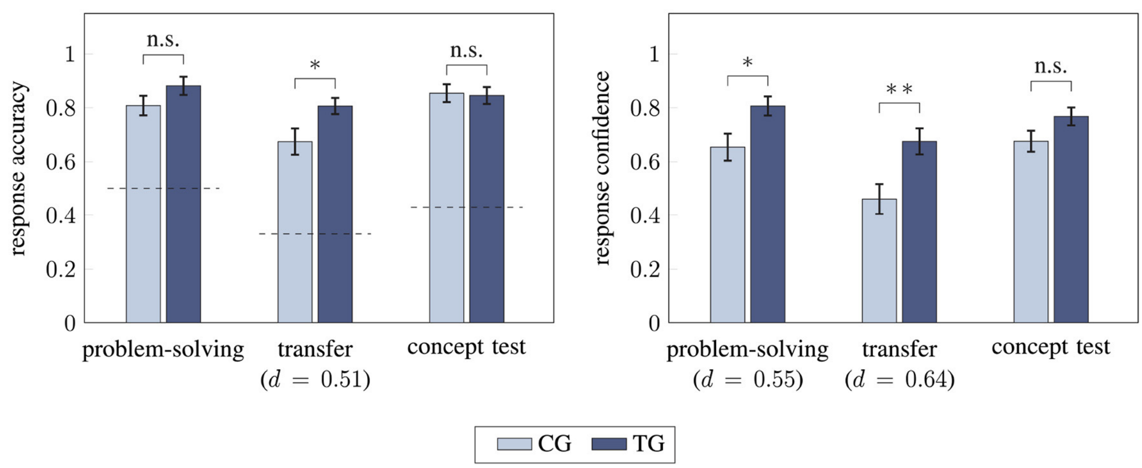


FIGURE 7
Comparison of students' performance in control and treatment group using unpaired (Welch) t-tests (two-sided). Response accuracy **(Left)** and response confidence **(Right)** for the problem-solving tasks, the transfer tasks, and the concept test are compared between treatment group (TG) and control group (CG; */** statistical significance ($p < 0.05$/$p < 0.01$), n.s. "(statistically) not significant" indicating p-values > 0.05, effect size d, the dashed lines indicate the guessing probability, error bars represent 1 SEM).

TABLE 3 Comparison of fixation count (FC), mean fixation duration (MFD in s), and time to first fixation (TFF in s) for the defined areas of interests (AOIs) on pre-exercises and divergence instruction between treatment group (TG) and control group (CG; mean, standard deviation, *p*-values of two-sided (Welch) *t*-tests for independent samples, n.s. "(statistically) not significant" indicating p-values > 0.05, effect size *d*; gaze data was recorded using mobile eye tracking).

AOI	Metric	TG	CG	*p*	*d*	TG	CG	*p*	*d*
Pre-exercise		Partial derivatives I				Partial derivatives II			
Instruction	FC	342.4 ± 143.2	202.3 ± 112.1	**0.005**	1.08	196.3 ± 81.7	125.4 ± 67.7	**0.013**	0.94
	MFD	0.59 ± 0.32	0.69 ± 0.60	n.s.	...	0.51 ± 0.28	0.51 ± 0.26	n.s.	...
	TFF	0.3 ± 0.6	0.1 ± 0.4	n.s.	...	0.5 ± 1.0	0.2 ± 0.6	n.s.	...
Task	FC	241.7 ± 104.6	81.0 ± 49.3	< **0.001***	1.93	154.5 ± 54.4	49.5 ± 28.3	< **0.001***	2.37
	MFD	0.57 ± 0.29	0.40 ± 0.19	n.s.	...	0.49 ± 0.25	0.43 ± 0.19	n.s.	...
	TFF	11.3 ± 10.9	16.7 ± 16.5	n.s.	...	1.3 ± 1.8	1.9 ± 6.7	n.s.	...
Information	FC	25.6 ± 13.9	30.1 ± 30.9	n.s.	...	34.4 ± 23.9	30.6 ± 27.0	n.s.	...
	MFD	0.79 ± 0.57	0.49 ± 0.26	n.s.*	...	0.54 ± 0.42	0.66 ± 0.58	n.s.	...
	TFF	4.9 ± 12.6	1.5 ± 2.3	n.s.	...	60.6 ± 40.9	34.9 ± 19.2	n.s.*	...
Divergence instruction									
Definition	FC	30.1 ± 18.9	30.7 ± 20.6	n.s.	...				
	MFD	0.66 ± 0.56	0.61 ± 0.47	n.s.	...				
	TFF	2.3 ± 5.4	2.3 ± 5.2	n.s.	...				
Strategy	FC	181.6 ± 91.1	121.1 ± 82.5	**0.059**	0.69				
	MFD	0.52 ± 0.41	0.45 ± 0.32	n.s.	...				
	TFF	18.9 ± 18.5	16.4 ± 8.7	n.s.*	...				
Diagram	FC	134.7 ± 74.2	79.6 ± 65.0	**0.034**	0.79				
	MFD	0.79 ± 0.47	0.85 ± 0.94	n.s.	...				
	TFF	10.7 ± 10.4	13.7 ± 12.8	n.s.	...				
Hint	FC	13.2 ± 10.4	16.0 ± 22.8	n.s.	...				
	MFD	0.50 ± 0.35	0.69 ± 0.55	n.s.	...				
	TFF	155.5 ± 98.0	115.4 ± 48.1	n.s.*	...				

* Welch *t*-test due to a lack of homogeneity of variance.

0.81 ± 0.40) reflecting small sized effects [$t(83) = -3.0$, $p = 0.004$, $d = 0.32$ and $\mathcal{X}^2(1) = 24.6$, $p < 0.001$, $OR = 1.098$; Figure 6]. After instruction, students' problem-solving and transfer scores were 0.85 ± 0.23 and 0.74 ± 0.26, respectively. Further, response confidence in the concept test increased significantly with large effect size [$t(83) = -8.4$, $p < 0.001$, $d = 0.92$].

While working with the instruction, students in the treatment and the control group answered the included questions equally [0.89 ± 0.18 and 0.90 ± 0.16 respectively, $p = 0.80$]. After instruction, the problem-solving score and the concept test score did not differ significantly between treatment and control group ($p = 0.14$ and $p = 0.85$) with both groups reaching performances above 80% (Figure 7; for example, for problem solving 0.88 ± 0.22 for the treatment group and 0.81 ± 0.23 for the control group). Students in the treatment group scored higher in the transfer tasks than those in the control group (0.81 ± 0.20 and 0.67 ± 0.31, respectively) and this difference was statistically significant with medium effect size [Welch $t(66.9) = 2.3$, $p = 0.02$, $d = 0.51$]. Additionally, students' confidence level in the treatment and the control group differed significantly in both the problem-solving and the transfer tasks, with medium effect sizes [Welch $t(72.5) = 2.5$, $p = 0.02$, $d = 0.55$ and $t(82) = 2.9$, $p = 0.004$, $d = 0.64$]. For the concept test, no such effect was found ($p = 0.08$). Additionally, there were no significant interaction effects between time and group for response accuracy or response confidence in the concept test ($p = 0.41$ and $p = 0.91$).

Furthermore, students in the treatment group reported significantly lower intrinsic cognitive load and higher germane cognitive load compared to those in the control group (Figure 8). The differences in intrinsic and germane cognitive load had small effect sizes [$t(82) = -2.2$, $p = 0.02$, $d = 0.47$ and $t(82) = 1.7$, $p = 0.0045$, $d = 0.37$]. No significant difference was found in extraneous cognitive load between the two groups ($p = 0.66$).

3.2. Study 2

3.2.1. Mobile eye-tracking analysis of pre-exercise and divergence instruction

Students fixation count on the vector field diagram of the initial problem VF_0 was 26.4 ± 19.5 with a mean fixation duration of (0.59 ± 0.39) s, which did not significantly differ between treatment and control group ($p = 0.91$ and $p = 0.91$). In average, students visited the first pre-exercise page (177.37 ± 73.15) s, the second pre-exercise page (122.5 ± 48.5) s, and the divergence instruction page (232.6 ± 93.5) s. Average visit durations differed significantly between the groups ($p < 0.001$, $p < 0.001$, and $p = 0.004$, respectively). More precisely, the treatment group visited all three pages significantly longer: the pre-exercise pages (383.1 ± 73.6) s [vs. (205.6 ± 33.4) s for the control group] and the divergence instruction page (275.4 ± 83.5) s [vs. (184.1 ± 81.5) s for the control group]. Results of detailed gaze analysis for the defined AOIs on the pre-exercise pages and the divergence instruction page are summarized in Table 3.

AOI-based analysis of visual behavior during completion of the pre-exercises on vector components and partial derivatives showed that students in the treatment group focused more frequently on the entire pre-exercise (AOI *Instruction*) and particularly on the drawing prompt (AOI *Task*). For both pre-exercise pages large differences were observed. These results suggest that students in the treatment group paid more attention to the task when they were asked to draw. Regarding the mean fixation duration and the time to first fixation on the defined AOIs, no group differences were found.

For the instruction on divergence, analysis revealed a significant between-subjects effect for the vector field diagram. It was more frequently fixated by students in the treatment group with medium effect size [$t(30.0) = 2.2$, $p = 0.034$ $d = 0.79$]. Further, the instructional strategy was more often fixated by the treatment group with medium effect size [$t(30.0) = 2.0$, $p = 0.059$, $d = 0.69$]. Neither the definition nor the hint were shown to be treated visually different by both groups. Regarding the mean fixation duration and the time to first fixation, no group differences were found. However, Table 3 indicates that students read the instruction from the top to the bottom, whereas the diagram was fixated slightly faster than the strategy by the treatment group [(18.9 ± 18.5) s strategy vs. (10.7 ± 10.4) s diagram] and the control group [(16.4 ± 8.7) s strategy vs. (13.7 ± 12.8) s diagram].

3.2.2. Mobile eye-tracking analysis VF_{1-4} and remote eye-tracking analysis of VF_{5-8}

In the first part of the assessment (mobile eye tracking, VF_{1-4}), the subjects' average total visit duration on the four task pages was (43.0 ± 13.6) s with no differences between the groups ($p = 0.20$). Additionally, the average fixation count of all subjects was 32.9 ± 18.9 with a mean fixation duration of (0.54 ± 0.34) s. Comparison of fixation-based metrics on the vector field diagram revealed no significant differences between treatment and control group (Table 4). A tendency for the control group to look at the diagram more frequently was observed, but it was not statistically significant.

TABLE 4 Comparison of fixation count (FC), mean fixation duration (MFD in s), and time to first fixation (TFF) on the vector field diagram for vector field plots VF_{1-8} and saccadic count (SC), proportion of horizontal and vertical saccades (SCP_H and SCP_V), and saccadic length of horizontal and vertical saccades (SL in pixels) on the vector field diagram for vector field plots VF_{5-8} between treatment group (TG) and control group (CG; mean, standard deviation, *p*-values of two-sided (Welch) *t*-tests for independent samples, n.s. "(statistically) not significant" indicating *p*-values > 0.05, effect size *d*).

Variable	TG	CG	*p*	*d*
VF_{1-4} *(mobile eye tracking)*				
FC*	28.1 ± 11.6	38.3 ± 24.0	n.s.	...
MFD	0.56 ± 0.32	0.51 ± 0.38	n.s.	...
TFF	1.7 ± 1.9	1.4 ± 1.7	n.s.	...
VF_{5-8} *(remote eye tracking)*				
FC*	38.9 ± 12.7	52.5 ± 19.9	0.002	0.82
MFD	0.27 ± 0.06	0.24 ± 0.06	n.s.	...
TFF	0.1 ± 0.3	0.09 ± 0.3	n.s.	...
SC	36.4 ± 12.3	49.7 ± 19.2	0.004	0.83
SCP_H	0.19 ± 0.07	0.21 ± 0.05	n.s.	...
SCP_V	0.14 ± 0.05	0.13 ± 0.05	n.s.	...
SL*/**	175.7 ± 140.9	192.8 ± 155.8	0.002	0.11

* Welch *t*-test due to a lack of homogeneity of variance.

** Reference: average distance between two adjacent vectors in the vector field diagram is 160 pixels.

In the second part of the assessment (computer-based remote eye tracking, VF_{5-8}), subjects spent in average (18.7 ± 7.1) s on the four task pages, without significant group differences ($p = 0.08$). Eye-tracking analysis for selected metrics on the vector field diagram are summarized in Table 4. By comparing treatment and control group, students in the treatment group fixated significantly less and performed significantly less saccades with large effect sizes [Welch $t(51) = -3.0$, $p = 0.002$, $d = 0.82$ and $t(51) = -3.0$, $p = 0.004$, $d = 0.83$]. In addition, mean fixation duration did not differ significantly between the groups, with the fixation duration varying between different fixations (Figure 9, Right). Further, there was a tendency for students in the control group to perform a higher percentage of horizontal saccades and for students in the treatment group to perform a higher percentage of vertical saccades than the respective other group; but no significant differences were found. The saccade plots (Figure 9, Left) indicate symmetrical gaze behavior along the coordinate directions for both groups. Oblique gaze directions were hardly found, regardless of group assignments. Last, students in the treatment group made significantly shorter horizontal and vertical saccades on the diagram with small effect size [Welch $t(2867.4) = -3.1$, $p = 0.002$, $d = 0.11$]. Average saccade length of the horizontal and vertical saccades on the diagram corresponds approximately to the distance between adjacent vectors (see Figure 9, Right). Regarding the time to first fixation, no group differences were found.

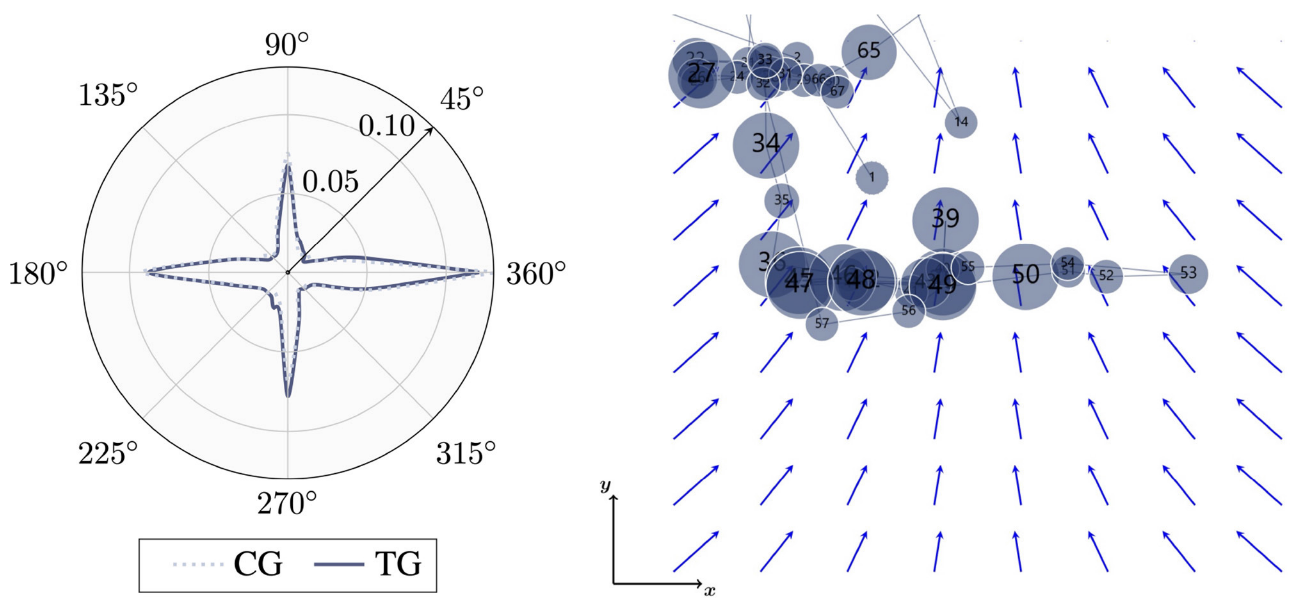


FIGURE 9
Polar distribution of saccade directions on the vector field diagrams VF_{5-8} for treatment group TG and control group CG **(Left)** and scanpath of a student from the treatment group (fixations are visualized by numbered circles; circle size relates to fixation duration) visually comparing adjacent vectors **(Right)**. Gaze data was recorded using remote eye tracking.

4. Discussion

In general, the instruction used in this study is based on prior work by Klein et al. (2018, 2019, 2021), which has been expanded upon theory and empirical findings by drawing activities and pre-exercises on vector decomposition and partial derivatives. Specifically, the prior work comprises eye-tracking studies that analyzed gaze behavior during instruction processing and subsequent problem-solving, which is also part of this study. Moreover, the study sample used in the prior studies and in this study are comparable, as all participants were first-year students at a German university with comparable prior knowledge. Due to the material development and a comparable study design and sample, in the following, the results of this study will be discussed in reference to the aforementioned prior work.

4.1. Study 1

4.1.1. Impact of pre-exercise and divergence instruction on students' achievement (RQ1)

Before instruction, students showed high prior knowledge regarding decomposition of vectors and a high conceptual understanding of divergence (Table 1). But only 39% of them were able to evaluate the divergence of a vector field diagram—a finding that was also reported by previous studies (Singh and Maries, 2013; Baily et al., 2016; Klein et al., 2018, 2019, 2021). After instruction with or without drawing activities, students' accuracy in judging a vector field's divergence and even their response accuracy and confidence in the concept test increased significantly. Then, 81% of the students evaluated the divergence of the initial vector field correctly. Moreover, they achieved a mean score of 85% for all eight vector fields in the problem-solving phase as well as for the concept test. These results underline the educational impact of the instruction, on the one hand, and going beyond, further emphasize the value of instructional support using multiple representations for complex physics concept, on the other hand.

Compared to previous, very similar studies, where students achieved scores of 69% after completing an integral and a differential instruction (Klein et al., 2018) and 77% after being instructed with or without visual cues (Klein et al., 2019), the instruction used here was able to contribute to an even higher improvement in judgment of a vector field's divergence. These findings particularly indicate the added value of the pre-exercises on vector decomposition and partial derivatives which support specific engagement with and learning of the two main concepts that are crucial for divergence. Following theories from cognitive psychology, by staggering the instruction as a step-by-step guide to the main explanation on divergence, the instruction is adapted to the limited working memory capacity (Baddeley, 1986; Rosenshine, 1995). In light of such an instruction as a step-by-step strategy, one could argue that students just learned how to follow the steps, but did not get a superordinate understanding of the underlying concepts—that are, vector decomposition and partial derivatives. However, the high transfer score (74%) referring to superordinate knowledge of partial derivatives beyond divergence, speaks against this effect and indicates an actual growth in conceptual understanding of students from both groups. This conclusion can be supported by comparing the findings with results from Klein et al. (2019), who referred to a matched transfer score of 54% when judging divergence and curl.

4.1.2. Impact of drawing activities on performance measures and cognitive load (RQ2, RQ3, H1)

The instructional support in the treatment group included drawing activities, such as highlighting rows and columns or sketching vector components, which was not provided in the control group. Comparing the results in the transfer tasks (loosely linked with the instruction) students from the treatment group performed particularly better and responded with higher confidence. The transfer tasks covered the concepts of covariation and partial derivatives, and went beyond mere step-by-step instructions on divergence. Our conclusion is that the drawing activities significantly enhanced students' understanding of vector fields and related concepts, such as covariation. Moreover, significant between-subject effects concerning students' perceived intrinsic and germane cognitive load indicated that the drawing activities supported and facilitated students' learning. As germane cognitive load refers to working memory resources that the learner devotes dealing with the matter to be learned, high values of germane cognitive load are associated with targeted devotion of working memory resources and optimized instructional procedures (Sweller, 2010). The aforementioned findings are in line with theories from cognitive psychology that promoted drawing as a powerful learning strategy in multi-representational learning environments by enhancing an effective use of working memory capacities (Bilda and Gero, 2005; Ainsworth et al., 2011; Kohnle et al., 2020; Ainsworth and Scheiter, 2021).

Comparisons between two groups during problem solving and in the concept test on divergence (close to instruction) showed that drawing activities did not yield to better performance, but students had more confidence in their answers. The latter is an equally positive outcome of effective teaching (Lindsey and Nagel, 2015; Klein et al., 2017, 2019). With students from the treatment group achieving a mean score of 88% during problem solving, drawing generated even higher learning outcomes then instructional support including visual cues (82% accuracy; Klein et al., 2019). Similar high performance scores during problem solving were also found for the control group (81%), indicating ceiling effects. These results imply that for tasks close to instruction, the impact of pre-exercises prevailed effects caused by the drawing activities. Here, further research regarding the actual impact of the pre-exercises by systematic manipulation of the instruction is required. Additionally, the positive correlation between performance and confidence that was found in previous studies also regarding instruction-based learning of divergence (Lindsey and Nagel, 2015; Klein et al., 2017, 2019) could be confirmed in this study.

4.2. Study 2

4.2.1. Visual processing of pre-exercises and divergence instruction (RQ4a)

During the instruction, students completed two pages of pre-exercises focused on vector decomposition and partial derivatives and one main instruction page on a visual interpretation of divergence. Since mean fixation duration did not differ between treatment and control group, it can be assumed that students invested the same cognitive effort for learning (Rayner, 1998; Ozcelik et al., 2009). Analysis of the average visit duration regarding the instructional material revealed significant group differences, i.e., students from the treatment group visited all three pages longer than students from the control group. More precisely, the treatment group spent almost twice as much time on the pre-exercise pages compared to the control group, while both groups scored high in the questions contained. Gaze differences were particularly found to result from more frequent fixations on the decomposition task stems. In this process, students systematically manipulated x or y components of vectors, either through drawing or mental imagination, to describe their changes along a row or column. Based on the eye-mind hypothesis, the drawing tasks facilitated a deeper processing of vector decomposition and the evaluation of their changes along coordinate directions, as represented by partial derivatives (Wu and Liu, 2022). Moreover, as both groups spent most of the time fixating the decomposition tasks, it can be assumed that pre-exercises provided a helpful foundation for the subsequent instruction on divergence.

In reference to the visual treatment of the divergence instruction, the findings from Klein et al. (2019) can be drawn upon, as the authors conducted eye-tracking analysis on a similar instruction that included and excluded visual cues. Average visit duration in the study by Klein et al. (2019, 106.8 ± 35.9 s) was similar to the control group's average visit duration in the study reported here, whereas the treatment group's average visit duration was found to be significantly higher. Moreover, definition of AOIs revealed that visual behavior on the diagram reflected the largest difference between students instructed with and without drawing activities—a result, that was also found by Klein et al. (2019, 2021). In general, while here fixation-based metrics for the AOIs *Diagram* and *Strategy* were higher compared to the results reported by Klein et al. (2019), visual attention on the AOIs *Definition* and *Hint* was less pronounced. In the study reported here, students fixated most on the AOIs *Strategy* and *Diagram*, whereas results for the control group in this study were mostly similar to the treatment group in the study by Klein et al. (2019). Thus, visual attention in this study was not evenly distributed over the whole divergence instruction page, but focused on the vector field diagram and the adjacent instructional text. This asymmetric distribution was particularly pronounced for the treatment group. By comparing the eye-tracking results of the previous study by Klein et al. (2019) with those of the study presented here, it thus appeared that drawing activities in the divergence instruction lead to even higher visual attention and, following the eye-mind hypothesis, deeper cognitive processing of the instruction's kernel than adding visual cues to the diagram (Wu and Liu, 2022). This is particularly remarkable since Klein et al. (2019) did not introduce a pre-exercise on component decomposition and partial derivatives which already prefigured central aspects of the divergence instruction. These findings are in line with previous results and theories from cognitive psychology reporting that drawing supported a deeper engagement with details and important parts within a learning environment (Hellenbrand et al., 2019; Kohnle et al., 2020; Ainsworth and Scheiter, 2021). This further underlines the educational impact of drawing activities. As such, eye-tracking data analysis during instruction-based learning allowed to explain the group differences regarding response confidence and cognitive load during problem solving (Section 4.1.2). However, it has to be noted that in the study by Klein et al. (2019) definition and hint were more elaborated than in the present study. Additionally, when comparing study results, the impact of the administration format needs to be taken into account, as in the studies by Klein et al. (2019, 2021) students completed the instruction on a computer screen and in the study reported here the instruction was given on physical paper sheets. Further, Klein et al. (2019, 2021) used remote eye-tracking, while in this study mobile gaze-data analysis was exploited (discussion see Section 4.3).

4.2.2. Gaze behavior during problem-solving (RQ4b)

In the problem-solving phase, visual behavior was analyzed using mobile as well as remote eye tracking. Comparing students visual behavior on VF_{1-4} with the initial problem VF_0 indicated that both instructions (with and without drawing activities) influenced students' visual handling of vector field diagrams. While no significant group differences for the first four tasks were found (discussion see Section 4.3), students from the treatment group performed significantly less fixations and saccades on the vector field diagram then students from the control group—a tendency that, however, was also indicated in the mobile gaze data. Since there were no differences in the mean fixation duration, this result also means that students from the treatment group needed less time for responding. In previous research, fewer fixations, and thus fast finding of a solution was shown to be associated with expertise (Reingold et al., 2001; Chi, 2006; Susac et al., 2014; Klein et al., 2018). In line with the theory of long-term working memory, which says that experts encode and retrieve information more rapidly than novices (Ericsson and Kintsch, 1995), fewer fixation counts are indicators of such rapid procedures (Gegenfurtner et al., 2011).

Moreover, students in the treatment group were shown to perform significantly shorter horizontal and vertical saccades on the vector field diagram. Since their mean length was close to the mean distance between two adjacent vectors, they exhibited behaviors that indicate a systematic comparison of adjacent vectors along coordinate directions (Figure 9, Right). These findings are in line with results in a previous study by Klein et al. (2018), where best-performing students were found to perform significantly shorter saccades compared to worst-performing students. Moreover, a study by Chen et al. (2014), which investigated students' gaze behavior during working on text- and picture based physics concept tasks, revealed that the mean saccade distance negatively predicts the success of retrieval performance in picture presentations, suggesting a greater probability that students will answer correctly if they make shorter saccade movements. Hence, the positive effect of drawing activities for learning of divergence which was found regarding different performance and cognitive load indicators (Section 4.1.2) and which is in line with theories from cognitive psychology, could be further supported by gaze-data analysis during problem solving.

Furthermore, the proportion of horizontal and vertical saccades on the vector field diagram, showed no significant differences between treatment and control group, however, both groups performed over 30% of horizontal and vertical saccades. For reference, given an equal distribution of gaze directions, approximately 11% of all gaze directions would be horizontal and vertical saccades (referring to a tolerance margin of $\pm5^\circ$). As a high proportion of horizontal and vertical saccades was shown to indicate a systematic handling of vector field plots by reflecting comparisons of vector components along the coordinate directions (Klein et al., 2018, 2019), these findings suggest expert-like procedures for students from both groups. Saccadic angle polar distribution (Figure 9, Left) supported this result as a symmetric distribution, where horizontal and vertical directions are pronounced whereas oblique directions are avoided, was associated with expertise in previous studies (Klein et al., 2018, 2019, 2021). While horizontal eye movements were found to be commonly dominant, for example, when looking at pictures, due to oculomotor eye factors and cultural reading habits, vertical eye movements are atypical and, therefore, can be associated with conscious problem-solving processes (Foulsham et al., 2008; Klein et al., 2021).

4.3. Methodological consideration: mobile and remote eye tracking in learning and assessment scenarios

Previous eye-tracking studies commonly used remote gaze data collection to analyze learning and assessment behavior by presenting work sheets, instructions, or tasks on a screen (Hahn and Klein, 2022a). However, in interactive settings, such as drawing-based learning, this method can no longer be applied. Then, mobile eye-tracking glasses provide the possibility to analyze cognitive processes by enabling a nearly natural setting and by not hindering the constructive process (Hellenbrand et al., 2019; Jarodzka et al., 2021). In contrast to typical methods used in such scenarios, for example, questionnaires and interviews, mobile gaze data analysis allows to gather data during actual behavior, thus increasing objectiveness of the measurement (Mayr et al., 2009). However, to the best of our knowledge, only few studies exploited mobile eye tracking in learning scenarios thus far (Hahn and Klein, 2022a). By investigating mobile eye gaze during learning with a drawing-based instruction, study 2 contributes to this line of research.

Further, by examining the same problem-solving task with both mobile and remote eye-tracking, first indications for a comparison of both eye-tracking methods in STEM, and particularly physics, education research assessment scenarios can be provided. Reconsidering the eye-tracking results from a methodological perspective, it was noticeable that students spent more than twice as much time per task in the mobile eye-tracking phase compared to the remote eye-tracking phase. In addition to learning- and routine-related effects, it seems reasonable to assume that this difference was also due to the method of eye-tracking data collection. Recording with glasses is unfamiliar, particularly for non-eyeglass wearers, and makes them more aware of the tracking (Mayr et al., 2009). This can lead to uncertainty and might influence students' behavior during assessment. Additionally, working on paper allows for active manipulation of the materials beyond the intended drawings, for example, by marking or taking notes, which are not available when the stimulus is given on a screen (Mayr et al., 2009). Furthermore, in the study presented here, it was noticeable that the glasses provide higher values for mean fixation duration and time to first fixation, while fixation count is significantly lower compared to the remote eye-tracking data when completing the same task. Consequently, mobile eye tracking showed no significant group differences regarding fixation-based metrics while remote eye-tracking did. Again besides learning and routine-related effects, this might be due to the glasses lower sampling rate and accuracy (Mayr et al., 2009). Particularly, the intermediate step of (manually or automatically) mapping video-recorded eye-tracking data to two-dimensional snapshots of the materials affects data accuracy. In case of manual mapping, fixations shown in the video are manually assigned to the corresponding location on the snapshot; saccades, in contrast, are automatically transferred without mapping (Tobii Pro Lab 1.204 software). Given the large number of fixations per video, this requires high time expenditure and leads to great error potential (e.g., skipping of fixations, inaccurate positioning). As a consequence, small effects that require a high level of measurement precision, for example, when analyzing mean fixation duration, or investigation of small AOIs with only a few fixations, can not be resolved precisely by the mobile gaze data.

Additionally, manually mapped gaze data and fixed AOIs only allow to illustrate processes of minimal interaction with the environment. However, to track processes in constructive settings, automatic adapting AOIs are required which are rarely available thus far (Wolf et al., 2018). Moreover, a look at the raw data of the eye-tracking glasses showed incomplete recordings of gaze paths as saccades, in particular, were mostly not completely tracked although the eye tracker indicated to have recorded a high percentage of gaze samples. However, as the calibration was system-controlled, no individual adjustment could be done. Additionally, accuracy is reduced as calibration refers to a fixation distance of 0.5–1 m which, however, continuously changes during data collection—a problem that was also reported to occur in other settings (Mayr et al., 2009). Particularly, analysis of saccadic angles and lengths is strongly affected by the described sources of error. To summarize the experiences in this study, mobile eye tracking provided additional insight into students' cognitive processes during drawing-based learning, however, the data needs to be evaluated and interpreted with caution. Particularly in scenarios that require detailed analysis of specific (particularly small) AOIs or saccade-based procedures, for example, navigating along coordinate directions of vector fields, mobile eye-tracking data is hardly resilient which should be kept in mind in light of the high time requirements for gaze-data mapping.

4.4. Conclusion and future work

In this work, the impact of drawing activities in multi-representational, instruction-based learning in the context of vector fields was investigated in two studies through analysis of different performance measures and eye-tracking data. Besides showing an immense overall impact of the instruction on students' conceptual understanding and their accuracy of judging a vector field's divergence, drawing activities were shown to led to significantly higher learning outcomes in the transfer task. Furthermore, intrinsic cognitive load of the learning subject was lower for students instructed with drawing activities, which increased their germane cognitive load enabling to devote more working memory resources in dealing with the subject matter to be learned. Moreover, eye-tracking analysis revealed that students instructed with drawing activities fixated important parts of the pre-exercises and the divergence instruction, that are, decomposition tasks, vector field diagram, and instructional text, more frequently. During subsequent assessment, both groups showed representation-specific, expert-like behaviors, such as comparing vectors in horizontal and vertical direction along the Cartesian coordinates, indicating a correct interpretation of partial derivatives. Furthermore, students instructed with drawing activities were found to be more effective compared to students instructed without drawing activities by fixating the vector field diagram less frequently in order to determine its divergence, and to systematically compare adjacent vectors along coordinate directions.

Concerning the value of this article for STEM education, it extends previous research on learning in the context of vector fields by showing how dedicated pre-exercises on vector decomposition and partial derivatives, on the one hand, and drawing activities, on the other hand, can further enhance previous instructions on divergence (Klein et al., 2018, 2019, 2021). An explicit focus on covariation for both groups, but particularly for those students who actively sketched them, not only supported students' visual knowledge of divergence, but was shown to actually deepened students' understanding of vector fields. For educational practice, the added value of a pre-exercise aimed at concrete practicing of vector decomposition and partial derivatives could be supported. However, since a systematic manipulation of the pre-exercise was not dedicated focus of this study, further research regarding its actual value and a meaningful design is required. To further continue this line of research, particularly regarding conceptual knowledge that could be transferred to associated concepts of vector fields, such as curl, huge potential emerge for combining drawing activities with other methods, for example, simulations (e.g., Kohnle et al., 2020; Ainsworth and Scheiter, 2021).

Although the instruction used here requires certain prior knowledge, and thus gears to university science students, some implications can also be transferred to other subjects and domains. For instructors, drawing activities aiming at step-by-step introducing a learning strategy or a problem-solving procedure can be recommended. Particularly, if most of the steps were usually done mentally or not explicitly introduced, drawing appears to be a promising learning method also beyond university learning, for example, in school.

From a methodological perspective, this article particularly benefited from existing prior work on previous divergence instructions and the same problem-solving task type (Klein et al., 2018, 2019, 2021), which enabled comparisons and conclusions regarding manipulations of the learning material and different methods of gaze-data collection. In this context, this article revealed valuable insights into mobile gaze-data collection and analysis in drawing-based learning and assessment scenarios, which were found to valuable complement performance and cognitive load data. However, although mobile eye tracking allowed to capture natural visual behaviors, particularly when the learner interacts with the learning environment, data obtained from eye-tracking glasses proved to be rather unsuitable for detailed analysis on a process level. In problem solving, a non-interactive setting, remote gaze data was shown to provide more reliable resolution of processes and strategies. By discussing limitations of mobile gaze-data analysis in educational settings and comparing mobile and remote eye tracking, this article may provide guidance and support for other researchers, who plan to study such cognitive processes with eye tracking. Specifically, a huge potential of mobile eye tracking emerges, for example, for collaborative learning and in experimental settings (e.g., Chien et al., 2015; Schneider et al., 2018), thus further development, for example, concerning supported mapping and automatic adaption of AOIs, is two-fold, required and promising.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

PK supervised data collection and gave feedback to the first draft of the manuscript. LH performed the statistical analyses and wrote the first draft of the manuscript. All authors contributed to conception and design of the study. All authors contributed to the article and approved the submitted version.

Funding

Students' expense allowance was funded by the Joachim Herz Foundation under Project No. 850030 (eyePhysics). Further, we acknowledge support by the Open Access Publication Funds of the Göttingen University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2023.1162281/full#supplementary-material

References

Ainsworth, S. E. (1999). The functions of multiple representations. *Comput. Educ.* 33, 131–152. doi: 10.1016/S0360-1315(99)00029-9

Ainsworth, S. E., Prain, V., and Tytler, R. (2011). Drawing to learn in science. *Science* 333, 1096–1097. doi: 10.1126/science.1204153

Ainsworth, S. E., and Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Curr. Direct. Psychol. Sci.* 30, 61–67. doi: 10.1177/0963721420979582

Alemdag, E., and Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Comput. Educ.* 125, 413–428. doi: 10.1016/j.compedu.2018.06.023

Ambrose, B. S. (2004). Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction. *Am. J. Phys.* 72, 453–459. doi: 10.1119/1.1648684

Baddeley, A. (1986). *Working Memory.* Oxford, UK: Oxford University Press.

Baily, C., Bollen, L., Pattie, A., Van Kampen, P., and De Cock, M. (2016). "Student thinking about the divergence and curl in mathematics and physics contexts," in *Proceedings of the Physics Education Research Conference 2016* (College Park, MD; New York, NY: American Institute of Physics), 51–54. doi: 10.1119/perc.2015.pr.008

Barniol, P., and Zavala, G. (2014). Test of understanding of vectors: a reliable multiple-choice vector concept test. *Phys. Rev. ST Phys. Educ. Res.* 10, 010121. doi: 10.1103/PhysRevSTPER.10.010121

Bilda, Z., and Gero, J. S. (2005). Does sketching off-load visuo-spatial working memory. *Stud. Desig.* 5, 145–160.

Bollen, L., Van Kampen, P., Baily, C., and De Cock, M. (2016). Qualitative investigation into students' use of divergence and curl in electromagnetism. *Phys. Rev. Phys. Educ. Res.* 12, 020134. doi: 10.1103/PhysRevPhysEducRes.12.020134

Bollen, L., Van Kampen, P., Baily, C., Kelly, M., and De Cock, M. (2017). Student difficulties regarding symbolic and graphical representations of vector fields. *Phys. Rev. Phys. Educ. Res.* 13, 020109. doi: 10.1103/PhysRevPhysEducRes.13.020109

Bollen, L., Van Kampen, P., and De Cock, M. (2015). Students' difficulties with vector calculus in electrodynamics. *Phys. Rev. ST Phys. Educ. Res.* 11, 020129. doi: 10.1103/PhysRevSTPER.11.020129

Bollen, L., Van Kampen, P., and De Cock, M. (2018). Development, implementation, and assessment of a guided-inquiry teaching-learning sequence on vector calculus in electrodynamics. *Phys. Rev. Phys. Educ. Res.* 14, 020115. doi: 10.1103/PhysRevPhysEducRes.14.020115

Burkholder, E. W., Murillo-Gonzalez, G., and Wieman, C. (2021). Importance of math prerequisites for performance in introductory physics. *Phys. Rev. Phys. Educ. Res.* 17, 010108. doi: 10.1103/PhysRevPhysEducRes.17.010108

Chen, S.-C., She, H.-C., Chuang, M.-H., Wu, J.-Y., Tsai, J.-L., and Jung, T.-P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Comput. Educ.* 74, 61–72. doi: 10.1016/j.compedu.2013.12.012

Chi, M. T. (2006). "Two approaches to the study of experts' characteristics," in *The Cambridge handbook of expertise and expert performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge, UK: Cambridge University Press), 21–30. doi: 10.1017/CBO9780511816796.002

Chien, K.-P., Tsai, C.-Y., Chen, H.-L., Chang, W.-H., and Chen, S. (2015). Learning differences and eye fixation patterns in virtual and physical science laboratories. *Comput. Educ.* 82, 191–201. doi: 10.1016/j.compedu.2014.11.023

Chiu, J. L., and Linn, M. C. (2014). Supporting knowledge integration in chemistry with a visualization-enhanced inquiry unit. *J. Sci. Educ. Technol.* 23, 37–58. doi: 10.1007/s10956-013-9449-5

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, M. P. (2006). Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. *Sci. Educ.* 90, 1073–1091. doi: 10.1002/sce.20164

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245. doi: 10.1037/0033-295X.102.2.211

Foulsham, T., Kingstone, A., and Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Res.* 48, 1777–1790. doi: 10.1016/j.visres.2008.05.018

Gegenfurtner, A., Lehtinen, E., and Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educ. Psychol. Rev.* 23, 523–552. doi: 10.1007/s10648-011-9174-7

Gire, E., and Price, E. (2012). "Graphical representations of vector functions in upper-division E&M," in *AIP Conference Proceedings* (New York, NY: American Institute of Physics), 27–30. doi: 10.1063/1.3679985

Griffiths, D. J. (2013). *Introduction to Electrodynamics, 4th Edn*. Boston, MA: Pearson.

Hahn, L., and Klein, P. (2021). "Multiple Repräsentationen als fachdidaktischer Zugang zum Satz von Gauß - Qualitative Zugänge zur Interpretation der Divergenz von Vektorfeldern," in P*hyDid B-Didaktik der Physik-Beiträge zur DPG-Frühjahrstagung*, eds J. Grebe-Ellis and H. Grötzebauch (Fachverband Didaktik der Physik, virtuelle DPG-Frühjahrstagung 2021), 95–100. Available online at: https://ojs.dpg-physik.de/index.php/phydid-b/article/view/1151/1237

Hahn, L., and Klein, P. (2022a). Eye tracking in physics education research: A systematic literature review. *Phys. Rev. Phys. Educ. Res.* 18, 013102. doi: 10.1103/PhysRevPhysEducRes.18.013102

Hahn, L., and Klein, P. (2022b). "Kognitive Entlastung durch Zeichenaktivitäten Eine empirische Untersuchung im Kontext der Vektoranalysis," in *Unsicherheit als Element von naturwissenschaftsbezogenen Bildungsprozessen*, eds S. Habig and H. van Vorst (Gesellschaft für Didaktik der Chemie und Physik, virtuelle Jahrestagung 2021), 384–387. Available online at: https://www.gdcp-ev.de/wp-content/tb2022/TB2022_384_Hahn.pdf

Hellenbrand, J., Mayer, R. E., Opfermann, M., Schmeck, A., and Leutner, D. (2019). How generative drawing affects the learning process: an eye-tracking analysis. *Appl. Cogn. Psychol.* 33, 1147–1164. doi: 10.1002/acp.3559

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learn. Instr.* 20, 172–176. doi: 10.1016/j.learninstruc.2009.02.013

Jarodzka, H., Skuballa, I., and Gruber, H. (2021). Eye-tracking in educational practice: investigating visual perception underlying teaching and learning in the classroom. *Educ. Psychol. Rev.* 33, 1–10. doi: 10.1007/s10648-020-09565-7

Jung, K., and Lee, G. (2012). Developing a tutorial to address student difficulties in learning curl: a link between qualitative and mathematical reasoning. *Can. J. Phys.* 90, 565–572. doi: 10.1139/p2012-054

Just, M. A., and Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cogn. Psychol.* 8, 441–480. doi: 10.1016/0010-0285(76)90015-3

Klein, P., Hahn, L., and Kuhn, J. (2021). Einfluss visueller Hilfen und räumlicher Fähigkeiten auf die graphische Interpretation von Vektorfeldern: Eine Eye-Tracking-Untersuchung. *Zeitschrift für Didaktik der Naturwissenschaften* 27, 181–201. doi: 10.1007/s40573-021-00133-2

Klein, P., Müller, A., and Kuhn, J. (2017). Assessment of representational competence in kinematics. *Phys. Rev. Phys. Educ. Res.* 13, 010132. doi: 10.1103/PhysRevPhysEducRes.13.010132

Klein, P., Viiri, J., and Kuhn, J. (2019). Visual cues improve students' understanding of divergence and curl: evidence from eye movements during reading and problem solving. *Phys. Rev. Phys. Educ. Res.* 15, 010126. doi: 10.1103/PhysRevPhysEducRes.15.010126

Klein, P., Viiri, J., Mozaffari, S., Dengel, A., and Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots? *Phys. Rev. Phys. Educ. Res.* 14, 010116. doi: 10.1103/PhysRevPhysEducRes.14.010116

Kohnle, A., Ainsworth, S. E., and Passante, G. (2020). Sketching to support visual learning with interactive tutorials. *Phys. Rev. Phys. Educ. Res.* 16, 020139. doi: 10.1103/PhysRevPhysEducRes.16.020139

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., et al. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* 10, 90–115. doi: 10.1016/j.edurev.2013.10.001

Leopold, C., and Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learn. Instr.* 22, 16–26. doi: 10.1016/j.learninstruc.2011.05.005

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., and Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* 45, 1058–1072. doi: 10.3758/s13428-013-0334-1

Li, J., and Singh, C. (2017). Investigating and improving introductory physics students' understanding of symmetry and Gauss's law. *Eur. J. Phys.* 39, 015702. doi: 10.1088/1361-6404/aa8d55

Lindsey, B. A., and Nagel, M. L. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. *Phys. Rev. ST Phys. Educ. Res.* 11, 020103. doi: 10.1103/PhysRevSTPER.11.020103

Logie, R. H. (2014). *Visuo-Spatial Working Memory*. London: Psychology Press. doi: 10.4324/9781315804743

Mayr, E., Knipfer, K., and Wessel, D. (2009). "In-sights into mobile learning: An exploration of mobile eye tracking methodology for learning in museums," in *Researching Mobile Learning: Frameworks, Tools and Research designs*, eds G. Vavoula, N. Pachler, and A. Kukulska-Hulme (Oxford: Peter Lang), 189–204.

Nieminen, P., Savinainen, A., and Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Phys. Rev. ST Phys. Educ. Res.* 8, 010123. doi: 10.1103/PhysRevSTPER.8.010123

Olsen, A. (2012). Tobii I-VT Fixation Filter. *Tobii Technol.* 21, 4–19.

Ozcelik, E., Karakus, T., Kursun, E., and Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Comput. Educ.* 53, 445–453. doi: 10.1016/j.compedu.2009.03.002

Pepper, R. E., Chasteen, S. V., Pollock, S. J., and Perkins, K. K. (2012). Observations on student difficulties with mathematics in upper-division electricity and magnetism. *Phys. Rev. ST Phys. Educ. Res.* 8, 010111. doi: 10.1103/PhysRevSTPER.8.010111

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educ. Psychol. Rev.* 29, 717–761. doi: 10.1007/s10648-016-9365-3

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372

Reingold, E. M., Charness, N., Pomplun, M., and Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychol. Sci.* 12, 48–55. doi: 10.1111/1467-9280.00309

Rosenshine, B. (1995). Advances in research on instruction. *J. Educ. Res.* 88, 262–268. doi: 10.1080/00220671.1995.9941309

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., and Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *Int. J. Comput. Support. Collab. Learn.* 13, 241–261. doi: 10.1007/s11412-018-9281-2

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learn. Instr.* 13, 227–237. doi: 10.1016/S0959-4752(02)00022-1

Shah, P., and Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *J. Exp. Psychol.* 125, 4–27. doi: 10.1037/0096-3445.125.1.4

Singh, C., and Maries, A. (2013). "Core graduate courses: a missed learning opportunity?," in *AIP Conference Proceedings* (New York, NY: American Institute of Physics), 382–385. doi: 10.1063/1.4789732

Smith, E. M. (2014). *Student* & *Textbook Presentation of Divergence [Master thesis]*. Corvallis, OR: Oregon State University.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., and Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educ. Stud. Math.* 104, 147–200. doi: 10.1007/s10649-020-09948-1

Susac, A., Bubic, A., Planinic, M., Movre, M., and Palmovic, M. (2019). Role of diagrams in problem solving: An evaluation of eye-tracking parameters as a measure of visual attention. *Phys. Rev. Phys. Educ. Res.* 15, 013101. doi: 10.1103/PhysRevPhysEducRes.15.013101

Susac, A. N., Bubic, A., Kaponja, J., Planinic, M., and Palmovic, M. (2014). Eye movements reveal students' strategies in simple equation solving. *Int. J. Sci. Math. Educ.* 12, 555–577. doi: 10.1007/s10763-014-9514-4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., and Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Comput. Educ.* 58, 375–385. doi: 10.1016/j.compedu.2011.07.012

Wolf, J., Hess, S., Bachmann, D., Lohmeyer, Q., and Meboldt, M. (2018). Automating areas of interest analysis in mobile eye tracking experiments based on machine learning. *J. Eye Mov. Res.* 11, 1–11. doi: 10.16910/jemr.11.6.6

Wu, C.-J., and Liu, C.-Y. (2022). Refined use of the eye-mind hypothesis for scientific argumentation using multiple representations. *Instr. Sci.* 50, 551–569. doi: 10.1007/s11251-022-09581-w

Wu, S. P. W., and Rau, M. A. (2018). Effectiveness and efficiency of adding drawing prompts to an interactive educational technology when learning with visual representations. *Learn. Instr.* 55, 93–104. doi: 10.1016/j.learninstruc.2017.09.010

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 31 July 2023
DOI 10.3389/feduc.2023.1141896

Check for updates

OPEN ACCESS

EDITED BY
Martin Rusek,
Charles University, Czechia

REVIEWED BY
Stanislav Popelka,
UPOL, Czechia
Sevil Akaygun,
Boğaziçi University, Türkiye

*CORRESPONDENCE
Ana Susac
✉ ana.susac@fer.hr

RECEIVED 10 January 2023
ACCEPTED 19 June 2023
PUBLISHED 31 July 2023

CITATION
Susac A, Planinic M, Bubic A, Jelicic K and Palmovic M (2023) Linking information from multiple representations: an eye-tracking study. *Front. Educ.* 8:1141896.
doi: 10.3389/feduc.2023.1141896

Linking information from multiple representations: an eye-tracking study

Ana Susac[1]*, Maja Planinic[2], Andreja Bubic[3], Katarina Jelicic[2] and Marijan Palmovic[4]

[1]Department of Applied Physics, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, [2]Department of Physics, Faculty of Science, University of Zagreb, Zagreb, Croatia, [3]Department of Psychology, Faculty of Humanities and Social Sciences, University of Split, Split, Croatia, [4]Laboratory for Psycholinguistic Research, Department of Speech and Language Pathology, Faculty for Special Education and Rehabilitation, University of Zagreb, Zagreb, Croatia

Eye tracking can provide valuable insights into how students use different representations to solve problems and can be a useful tool for measuring the integration of information from multiple representations. In this study, we measured the eye movements of 60 university students while solving two PISA items that contain graphs taken from mathematics and science assessments with the aim of studying the difference in visual attention between students who correctly and incorrectly identify graphs from a verbal description. We were particularly interested in the differences in the integration of information from different representations (text, graphs, and picture) between students who were successful or unsuccessful in solving items. The results suggest that students who solved the items correctly tend to solve the items longer than their counterparts who did not solve the items correctly. Analysis of eye tracking data suggests that students who solved science item correctly analyzed the graph for significantly longer time and had significantly longer average fixation time. This finding suggests that a careful analysis of graphs is crucial for the correct solution of PISA items used in this study. Furthermore, the results showed that students who solved the mathematics item correctly had significantly higher number of transitions between graphs and picture, which indicates a greater integration of information from two different representations. This indicates that these types of items require a lot of time and effort to complete, probably because solving them requires a lot of steps, which is cognitively demanding. We also found that the average fixation durations for different representations may vary for different items, indicating that it is not always equally difficult to extract necessary information from different types of representations. The results of this study suggest that instructors may be able to improve their teaching methods by considering the importance of individual representations (e.g., texts, graphs, and pictures) and the integration of information from multiple sources.

KEYWORDS
eye tracking, STEM education, visual attention, multiple representations, graphs, PISA

1. Introduction

There is a wealth of research in the field of education that supports the idea that using multiple representations in teaching and learning can be beneficial for students. One early theory that contributed to this understanding is Paivio's dual coding theory (Paivio, 1971), which assumes that verbal and pictorial information are processed in distinct cognitive systems. This was confirmed by functional brain imaging studies and led to the development of the well-known multimedia learning theory (Mayer, 1997). This theory suggests that presenting information in different representations (e.g., words and pictures) is more advantageous for learning than using single representations (e.g., words alone).

Another theory that helps to explain the benefits of using multiple representations is the cognitive load theory (Sweller, 1988). This theory suggests that students have a limited capacity for processing information in their working memory and that presenting information in multiple representations can help to reduce the cognitive load and make understanding easier. For example, presenting the same information using multiple representations can reduce cognitive load by making it easier for learners to extract the key information and understand the relationships between different concepts. Namely, a graph can provide a visual representation of the same information that was presented in the text, making it easier for learners to understand and remember.

However, it is important to consider that excess information that does not scaffold the learner can have a negative effect on learning. When developing instructional methods and materials, it is important to carefully consider the use of multiple representations. For example, research has shown that the inclusion of pictures in a text can be helpful for students in some cases, but it can also have negative effects on learning (Schnotz and Bannert, 2003). The use of multiple representations can be an effective way to engage students and facilitate learning, but it is crucial to consider the specific needs and abilities of the learners and to use them in a targeted and effective manner. Some researchers developed conceptual frameworks for learning with multiple representations based on constructivist theories of education and research findings on learning and teaching with different representational formats (Ainsworth, 2006; Airy and Linder, 2017).

Eye tracking is a research method that is commonly used in education research to study how people learn and process information. In recent years, there have been numerous eye-tracking studies conducted in the field of physics education, and a review by Hahn and Klein (2022) provides an overview of the main topics and findings from these studies. For example, eye-tracking was used to investigate students' visual attention while taking standard tests such as the Force Concept Inventory (FCI; Han et al., 2017; Kekule and Viiri, 2018). Küchemann et al. (2020) showed that eye tracking can be used for the identification of preconceptions related to rotating frames of reference. Different strategies in the interpretation of the divergence of the graphical vector field were also explored by using eye tracking (Klein et al., 2018, 2019c). Hoyer and Girwidz (2020) compared the eye movements of students to assess the effect of animation and interactivity in a computer-based physics experiment.

One area that has been explored through eye-tracking research is how students use different representations during problem solving. It has been shown that graphical representations of measurement data can be particularly helpful for students, as they can assist with visualizing and understanding the data (Susac et al., 2017). Supportive diagrams that visualize the physical situation in physics problems can also be beneficial, as they can reduce cognitive load and free up cognitive resources for further problem solving (Susac et al., 2019).

Pictorial representations have also been found to be effective in conveying physics concepts. Chen et al. (2014) found that it is easier and faster to identify crucial areas in a picture than in text while Chen and She (2020) found that pictorial representation helped students to better understand electricity concepts compared to a textual representation. However, it is worth noting that even pictorial representations can contain complex information that is not always easy to understand, as was demonstrated in a study on the recognition of pictorial representation of interference and diffraction patterns in wave optics (Susac et al., 2020, 2021). These studies found that this was a very demanding item for students, not a mere recall of a remembered pattern, thus indicating that pictorial representations can sometimes be challenging to grasp.

Eye tracking has been used extensively to investigate student understanding of graphs in physics. For example, an early study investigated the link between spatial visualization ability and solving kinematics problems and interpreting kinematics graphs (Kozhevnikov et al., 2007). Madsen et al. first analyzed how visual attention differed between those who answered correctly and incorrectly introductory physics problems (some of which contained graphs), and in the subsequent study, they investigated the effects of visual cueing on students' eye movements and performance on similar problems (Madsen et al., 2012, 2013). Kekule (2014) also explored the differences in the visual attention of students who solved correctly and incorrectly questions with kinematics graphs. Several studies examined students' eye movements while they were solving questions probing their understanding of graph slope and area under a graph (Susac et al., 2018; Klein et al., 2019a,b; Brückner et al., 2020). Furthermore, the identification of graphs that describe certain physical phenomena was investigated for students who were divided by their physics teacher into one of three groups according to their success in physics classes (Skrabankova et al., 2020). A recent study explored how students extract information from complex graphical displays of information such as the Hertzsprung-Russell diagram (Langendorf et al., 2022).

Research has found that the processing of information conveyed through graphs is complex and can take a significant amount of time to extract the necessary information (e.g., Susac et al., 2018; Klein et al., 2019a). Eye-tracking studies have also shown that experts and non-experts have different strategies for solving items that contain graphs and that students generally struggle with interpreting and analyzing graphs (e.g., Madsen et al., 2012; Susac et al., 2018; Klein et al., 2019a). One area where students have particularly been found to have difficulties is in conceptual understanding and calculation of the area under the graph (Susac et al., 2018; Klein et al., 2019a). Improving students' understanding and ability to interpret and analyze graphs is an important area of focus in physics education, and the use of eye tracking can help researchers and educators better understand how students approach these items and how they can be supported in their learning.

In addition to examining the use of individual representations, such as texts, graphs, pictures, equations, etc., and comparing their effectiveness, it is also valuable to investigate the use of multiple representations in STEM teaching and learning more broadly. Research in this area can help educators understand how to effectively integrate different representations to support student learning. Eye tracking can be a useful tool for studying the integration of information from different representations, as it allows researchers to track how students use and process multiple forms of information while solving problems (Rau, 2017). Multiple representations can positively affect learning by providing learners with different ways to approach and understand concepts, but too many representations can also lead to confusion and hinder the learning process.

In a recent study, Wu and Liu (2021) found that students with higher prior knowledge had a greater number of eye-movement transitions between representations compared to those with lower prior knowledge. Van Gog et al. (2005) reported some expertise-related differences in electrical circuit-troubleshooting performance while Kekule and Viiri (2018) found differences in the way that students who solved items correctly and incorrectly used different representations. These differences depended on the specific type of representation and the item at hand. For example, students who correctly solved items involving graph representations tended to look at the entire area of the graphs, while those working with motion map representations focused on individual points. Motion map representations depict the motion of an object over time, with each mark on a horizontal line indicating the position of the object at a specific point in time. Ibrahim and Ding (2021) also found that the integration of information from a diagram and text depended on the type of problem being addressed. Overall, the use of multiple representations in physics teaching and learning can be beneficial, and eye tracking can provide valuable insights into how students process and integrate information from different forms of representation.

Multiple representations have also been investigated in other STEM education disciplines such as mathematics and chemistry. For example, Ott et al. (2018) studied the use of different combinations of representations in mathematics problem solving and found that the combination of text and formula was as effective as other combinations containing more representations. In this study, text representation was found to be the most attended to and can be regarded as the reference representation.

Similarly, Stieff et al. (2011) found that students struggle with multiple representations in molecular mechanics and tend to attend more to visual–spatial representations (ball-and-stick model of the molecular system) than mathematical representations (equations). O'Keefe et al. (2014) also explored the integration of information from multiple representations in a multimedia simulation of the ideal gas law and found that transitions between different simulation elements were related to different learning outcomes. The authors emphasized the importance of making conceptual connections between specific representations in the learning process.

In this study, we decided to use Program for International Student Assessment (PISA) items that contain graphs because graphs are often used in PISA items to present information about scientific phenomena or to prompt students to interpret and analyze data. Thus, students need to extract and integrate information from text and graphs to solve the items. In addition, PISA items are designed to evaluate the general knowledge and skills of 15-year-old students in different countries, i.e., they measure "the ability to complete tasks relating to real life, depending on a broad understanding of key concepts, rather than limiting the assessment to the understanding of subject-specific knowledge" (OECD, 2007).

The focus on general knowledge and skills in PISA items is intended to measure students' ability to think critically and creatively about scientific issues and to use scientific knowledge and skills to solve problems. These are important skills for success in higher education, and PISA items provide a useful way to assess students' progress in these areas. While the items are designed specifically for 15-year-old students, it is expected that older students should also be able to solve these general items, as they do not require very specific knowledge. The OECD (2018) emphasizes the importance of these skills and the value of PISA items in measuring them.

There have been several previous eye-tracking studies that have used PISA items. For example, Krstić et al. (2018) analyzed the eye movements of 15-year-old students while they were solving PISA reading items. Hu et al. (2017) investigated how high-performing and low-performing students solve different types of PISA problems. Tóthová and Rusek (2022) compared how chemistry students and chemistry experts solve chemistry and general PISA science items and found that the experts were more efficient, needed less time, and focused on relevant parts of the items more than the students. Thomaneck et al. (2022) used PISA mathematics items in a study on the use of the eye-mind hypothesis in the domain of functions. Lundgren (2022) developed a computational model that simulates different strategies for solving a PISA problem-solving item. The study showed that simulations can be useful in understanding how changing a problem's properties affects our ability to infer problem-solving strategies.

These studies demonstrate the usefulness of PISA items in eye-tracking research, as they provide a standard way to measure student knowledge and skills and allow for comparisons between different groups of students or experts. PISA items are widely used in education research, and the use of eye tracking in studying these items can provide valuable insights into how students attend to and process different forms of information.

In this study, we aim to answer the following research questions:

RQ1. What is the difference in the visual attention between students who correctly and incorrectly identify graphs from a verbal description?

Visual attention refers to the extent to which participants focus on a specific representation (text, graphs, or picture) and can be operationalized through eye-tracking measures, such as dwell time and average fixation duration.

RQ2. What is the difference in the integration of information from multiple representations between students who correctly and incorrectly answer questions?

Integration of information refers to the extent to which participants are able to combine and put together information from multiple representations (such as text, graphs, or picture) to answer a question and can be operationalized through the number of transitions between representations. Participants who correctly answer questions are expected to show greater integration of information from multiple representations than those who do not.

2. Materials and methods

2.1. Participants

Participants in this study were 60 (34 female and 26 male) undergraduate university students in different years of study. They had diverse backgrounds (science, engineering, humanities, etc.) and their mean age was (23 ± 3) years. We used convenient sampling; voluntary participants who were prepared to come to the university for eye-tracking measurement and to answer some mathematics and physics questions. All participants gave informed written consent before taking part in the study.

2.2. Materials

Students answered nine PISA mathematics and science items that were released in 2006 (OECD, 2006a,b) and that were selected so that each item contained a graph in question stem or in multiple-choice options. Since the goal of this study was to investigate the differences in visual attention and integration of information from multiple representations between students who correctly and incorrectly answer questions, we will report results on only two PISA items Q1 and Q2. For these two items, students' scores were 52 and 48%, i.e., the numbers of students who answered correctly and incorrectly were comparable. For other PISA items, students' scores were considerably higher, so the numbers of students who answered correctly and incorrectly were not comparable, so we did not analyze them further.

In PISA science item Q1 (S529Q02; OECD, 2006b), four working conditions of electricity generation in a wind farm are described and students are asked which of the given graphs best represents the relationship between wind speed and electric power output (Figure 1A). In PISA mathematics item Q2 (M465Q01; OECD, 2006a), it is described that water is poured into the water tank whose picture is shown. Students should answer which of the given graphs shows how the height of the water surface changes over time (Figure 1B).

2.3. Procedure

Eye movements were recorded using the SMI iView Hi-Speed system with a sample rate of 500 Hz and the SMI screen-based RED-m system with a sample rate of 120 Hz (SensoMotoric Instruments G.m.b.H.). The eye-tracking system was calibrated for each participant before the data recording using a 13-point calibration algorithm. Questions were presented on a monitor at a distance of 50 cm from the participants' eyes. By choosing the answer, participants advanced to the next question. There was no time limit to answer the questions.

After the measurement of eye movements, students solved the same questions using a paper-and-pencil test and gave explanations for their answers. We asked the participants to provide an explanation afterward to make sure they did not choose the correct answer by chance or for the wrong reason. The whole procedure, including eye-movement calibration, recording, and paper-and-pencil testing lasted around 40 min.

2.4. Data analysis

Students' responses to the nine PISA items were scored correct or incorrect. In addition, these scores were corrected, considering students' answers and explanations in the paper-and-pencil test. If a correct answer during the eye-tracking measurement was given with a correct explanation in the paper-and-pencil test, the student was awarded one point. If a correct answer was given with a wrong explanation, the student was awarded 0 points. A correct answer without a correct explanation indicated that the correct answer was probably selected by chance or for a wrong reason. The correction of students' responses given during the eye-tracking measurement was rare, it happened in only 2.4% of all questions. Students' scores after the correction are reported in this paper.

The recorded eye movements data were analyzed using BeGaze software which allows evaluation of the eye fixations and saccades. Fixation is the state in which the eye is stationary over a period of time, while saccade is the rapid eye movement between fixations.

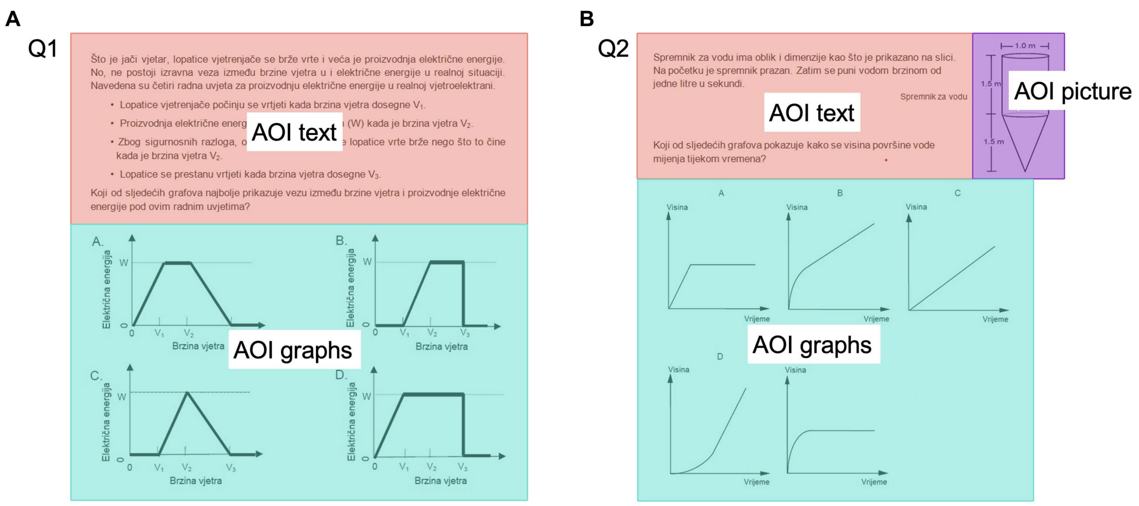


FIGURE 1
(A) Definition of areas of interest (AOIs) for item Q1. **(B)** Definition of areas of interest (AOIs) for item Q2.

BeGaze used the identification by dispersion-threshold (IDT) algorithm to determine fixations with a maximum dispersion value 100 px and minimum fixation duration 80 ms.

The appropriate number of areas of interest (AOIs) was defined for each item. Figure 1 shows AOIs for items Q1 and Q2. AOIs text included the introduction text and question, and AOIs graphs multiple-choice answers given as graphical representations. Item Q2 also contained an AOI picture, a pictorial representation of a water tank. We evaluated the dwell time, the number of fixations, the number of revisits, and the average fixation duration for each AOI. As we previously reported, these eye-tracking measures are dependent, and usually show a similar pattern of responses (Susac et al., 2019); thus, we will report results on dwell time and average fixation duration. In addition, we evaluated transitions between AOIs as a measure of the integration of information from multiple representations.

Reaction time refers to the amount of time it takes for a person to respond to a PISA item. Dwell time starts at the moment the AOI is fixated and ends at the moment the last fixation on the AOI ends. Fixation duration refers to the average duration of single eye fixations and it typically ranges from 100 to 600 ms (Hahn and Klein, 2022).

Student's t-test and several two-way ANOVAs were conducted in the analysis of eye-tracking data. A threshold of $p = 0.05$ was used for determining the level of effect significance within all conducted tests.

3. Results

3.1. Linking information from text and graphs (Q1)

Students who correctly answered item Q1 had the mean reaction time (RT) and standard deviation (80 ± 22) s whereas their peers who incorrectly answered the same item needed (68 ± 21) s to respond. The difference was statistically significant ($t(58) = 2.18$, $p = 0.03$).

To compare the distribution of visual attention of students who answered item Q1 correctly and incorrectly, we conducted two-way ANOVAs with repeated measures on factor AOI (text vs. graphs), while the between-subjects factor was Group (correct vs. incorrect). For dwell time, the results showed a statistically significant main effect of both factors, AOI [$F(1,58) = 150.41$, $p < 0.0001$, $\eta_p^2 = 0.72$] and Group [$F(1,58) = 5.79$, $p = 0.02$, $\eta_p^2 = 0.09$], whereas interaction effect was not statistically significant [$F(1,58) = 0.15$, $p > 0.05$, $\eta_p^2 = 0.003$]. Students had a significantly longer dwell time for AOI text [(47 ± 16) s] than AOI graphs [(22 ± 9) s]. Students who correctly answered the question had mean dwell time (37 ± 15) s that was significantly longer than the dwell time (31 ± 13) s of their peers who gave an incorrect answer (Figure 2A). In particular, the dwell time for AOI graphs was (24 ± 9) s for correct solvers, and it was significantly longer than (19 ± 8) s for incorrect solvers [$t(58) = 2.38$, $p = 0.02$].

For average fixation duration, a significant main effect of AOI [$F(1,58) = 6.15$, $p = 0.02$, $\eta_p^2 = 0.10$] was found, whereas the effects of Group [$F(1,58) = 0.06$, $p > 0.05$, $\eta_p^2 = 0.001$] and interaction effect [$F(1,58) = 0.01$, $p > 0.05$, $\eta_p^2 = 0.0002$] were not significant. Figure 2B illustrates that the average fixation duration was significantly longer for AOI graphs [(235 ± 33) ms] than for AOI text [(223 ± 43) ms].

To quantify the integration of information from text and graphs, we evaluated the number of transitions between AOI text and AOI graphs (Figure 3). The mean number of transitions and standard deviation for students who answered item Q1 correctly was 20 ± 11, compared to 17 ± 14 for students who answered incorrectly. The difference was not statistically significant [$t(58) = 0.67$, $p > 0.05$].

Furthermore, we created a sequence chart to visualize the distribution of fixations on AOI text and AOI graphs for students who correctly and incorrectly answered item Q1 (Figure 4). The sequences of eye movements show how often and how long the students attended each AOI. Although Figure 4 shows differences in the visual attention of the participants while they were solving item Q1, they mostly read the text of the task first and then looked at the graphs, occasionally returning to the text. Some participants switched attention from one AOI to another frequently, while others had a much smaller number of transitions between the AOIs.

3.2. Linking information from text, picture, and graphs (Q2)

Mean RT and standard deviation were (51 ± 23) s for students who correctly answered item Q2 and (43 ± 15) s for students who incorrectly answered the same item. The difference was not statistically significant [$t(58) = 1.63$, $p > 0.05$].

To compare the students' dwell time and average fixation duration, we conducted a two-way mixed design ANOVA with a between-subjects factor Group (correct vs. incorrect) and within-subjects factor AOI (text vs. graphs vs. picture). For dwell time, we found a significant main effect of AOI [$F(2,116) = 44.46$, $p < 0.0001$, $\eta_p^2 = 0.43$], whereas the effect of Group [$F(1,58) = 2.79$, $p > 0.05$, $\eta_p^2 = 0.04$] and interaction effect [$F(2,116) = 2.87$, $p > 0.05$, $\eta_p^2 = 0.05$] were not significant. *A priori* planned comparison of dwell time for two groups (correct and incorrect) on each AOI with Bonferroni-corrected p-values revealed statistically significant difference only for AOI graphs [$t(58) = 2.76$, $p = 0.02$]. Students who correctly answered item Q2 had dwell time (18 ± 9) s for AOI graphs that was significantly longer than the dwell time (13 ± 5) s of students who answered incorrectly (Figure 5A).

For average fixation duration, a significant main effect of AOI [$F(2,116) = 47.32$, $p < 0.0001$, $\eta_p^2 = 0.45$] was found, whereas the effect of Group [$F(1,58) = 3.76$, $p = 0.06$, $\eta_p^2 = 0.06$] and interaction effect [$F(2,116) = 0.16$, $p > 0.05$, $\eta_p^2 = 0.003$] were not significant. Figure 5B indicates that the average fixation duration was the longest for the AOI picture. Although Figure 5B shows a trend that students who answered the question correctly have a longer average fixation time, no *a priori* planned comparison of average fixation time for two groups on each AOI with Bonferroni-corrected p values reached a statistically significant difference. In fact, if we did not correct the p-values for multiple comparisons, we would get the outcome that students who correctly answered question Q2 had a longer average fixation duration for AOI graphs than their peers who failed to do so. Their mean fixation durations were (245 ± 47) and (221 ± 35) ms, respectively, and they were not statistically significantly different.

Again, we used a number of transitions as a measure of the integration of information from text, graphs, and picture. Figure 6 shows the number of transitions between text and picture, text and graphs, and graphs and picture. A two-way mixed-design ANOVA was performed with a between-subjects factor Group (correct vs. incorrect) and within-subjects factor Type of transition (text ↔ picture vs. text ↔ graphs vs. graphs ↔ picture). The results revealed a statistically significant main effect of Type of transition

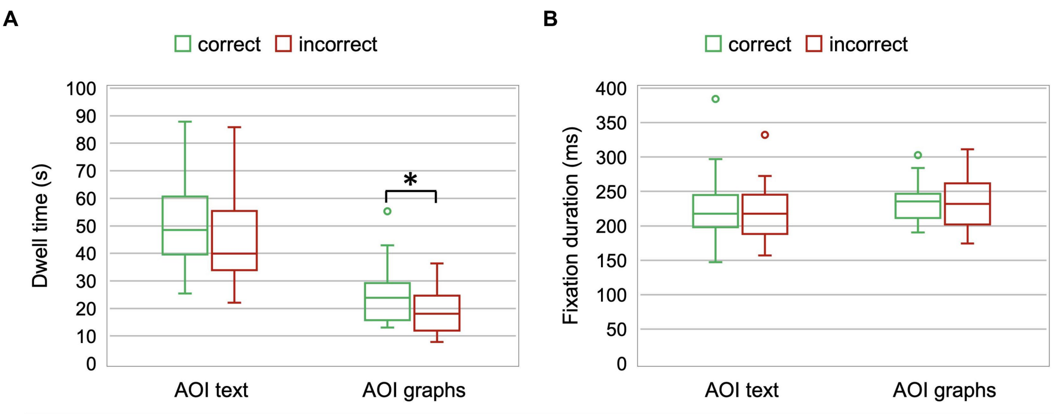


FIGURE 2
(A) Dwell time of students who correctly and incorrectly answered item Q1 for AOI text and AOI graphs. A box and whisker chart shows median, distribution of data into quartiles, and outliers. **(B)** Average fixation duration of students who correctly and incorrectly answered item Q1 for AOI text and AOI graphs.

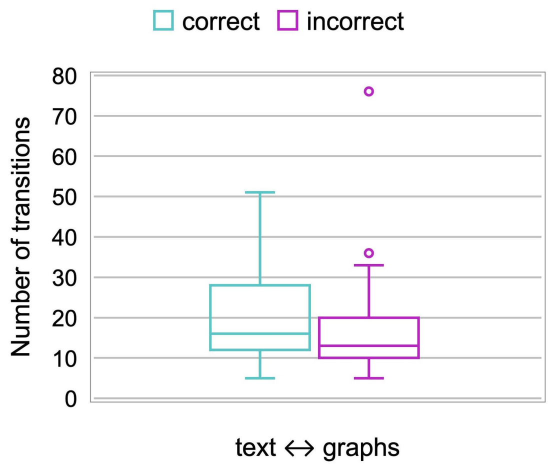


FIGURE 3
Number of transitions between AOI text and AOI graphs of students who correctly and incorrectly answered item Q1.

[$F(2,116) = 22.99$, $p < 0.0001$, $\eta_p^2 = 0.28$] while the effect of Group did not reach statistical significance [$F(1,58) = 3.92$, $p = 0.05$, $\eta_p^2 = 0.06$]. The interaction effect of these two factors was statistically significant [$F(2,116) = 4.65$, $p = 0.01$, $\eta_p^2 = 0.07$]. *A priori* planned comparison of the number of transitions for two groups (correct and incorrect) on each pair of AOIs with Bonferroni-corrected *p*-values showed statistically significant difference only for transitions between AOI graphs and AOI picture [$t(58) = 2.74$, $p = 0.02$]. Students who correctly answered item Q2 had a significantly larger number of transitions graphs ↔ picture than their peers who answered incorrectly (12 ± 9 and 7 ± 6, respectively).

To visualize the distribution of fixations at particular AOIs in time, we created a sequence chart. Figure 7 shows that students mostly first read the text, occasionally looking at the AOI picture, and then mostly looked at the graphs, sometimes returning to the AOI picture and/or AOI text. There is also a trend that students who correctly solved item Q2 spent more time paying attention to the AOI graphs than their peers who did not solve the task correctly.

4. Discussion

The results of this study suggest that students who correctly solved the two PISA items containing graphs took a longer time to do so than their peers who did not give correct answers. This indicates that some complex PISA items require a longer time to be understood and solved. This is not very surprising, considering that it is unlikely, given the usual types of tasks related to graphs used in mathematics and physics teaching in Croatia, that students are familiar with PISA kind of problems, and it is generally expected that it will take longer to solve unfamiliar problems than familiar ones. Similar results were obtained by Tóthová and Rusek (2022) who found that the student who was successful in solving the PISA item took the longest to do so.

To investigate this further, we compared the dwell times of students who correctly and incorrectly solved items for different representations (text, graph, and picture). For item Q1, which contained text and graphs, students who solved the item correctly had a longer dwell time for AOI graphs. In this item, the text of the task is quite long, and the students attended the text more than the graphs. Item Q2 contained text, graphs and a picture and students spent the least amount of time looking at the picture. This may be because the picture included relatively less relevant information for completing the task, which could have led students to spend less time on it compared to the text and graphs. The only statistically significant difference between the dwell times of students who correctly and incorrectly solved the item was found for AOI graphs. Students who solved the item correctly analyzed the graph for a longer time. This suggests that graph analysis is essential for the correct solution of this item.

Furthermore, different average fixation duration for different representations indicate that it is not always equally difficult to extract the necessary information from text, graphs, and pictures. For question Q1, the average fixation duration for AOI graphs was longer compared to the average fixation duration for AOI text. On the other hand, for question Q2, the average fixation duration was the longest for AOI picture. Again, a trend was shown that the difference between the two groups of students (those who solved the item correctly and incorrectly) appeared for AOI graphs. Again, the key to solving problems correctly seems to be the ability to extract relevant

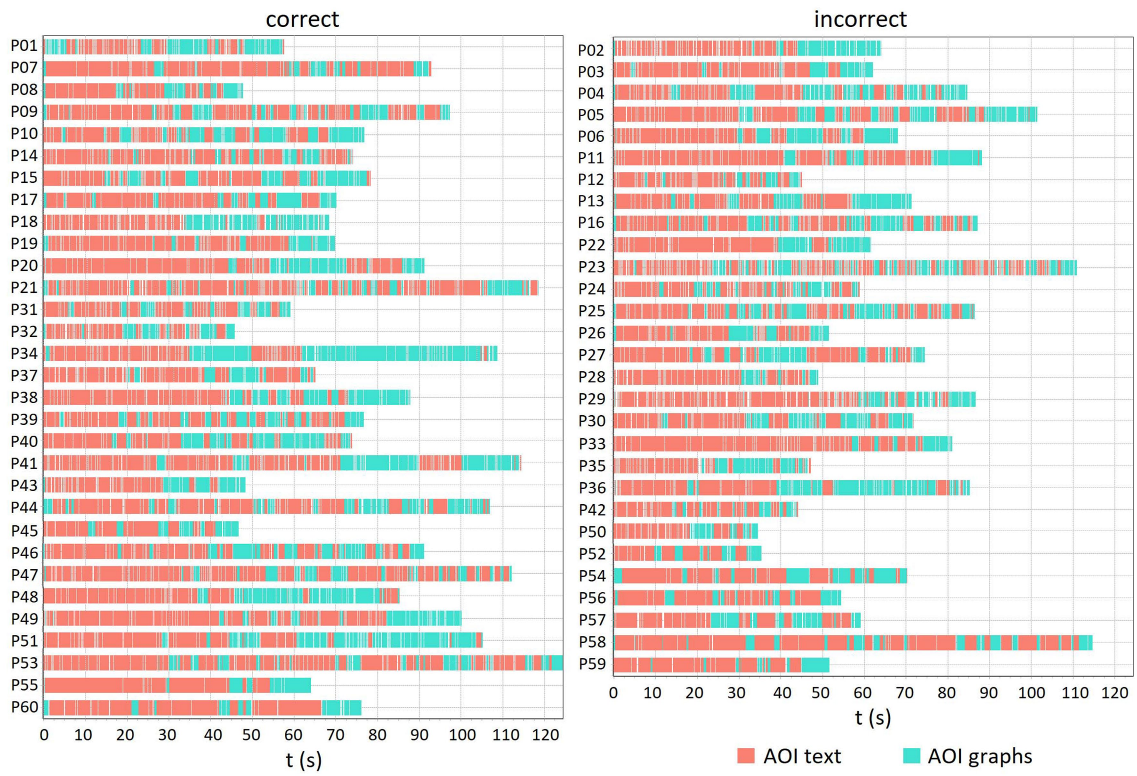


FIGURE 4
The sequence chart for PISA item Q1 shows the order of fixations for AOI text and AOI graphs, separately for students who solved the item correctly and incorrectly. Fixations that fell within the boundaries of the AOIs are color-coded based on the color of the AOI. Each row represents a different participant.

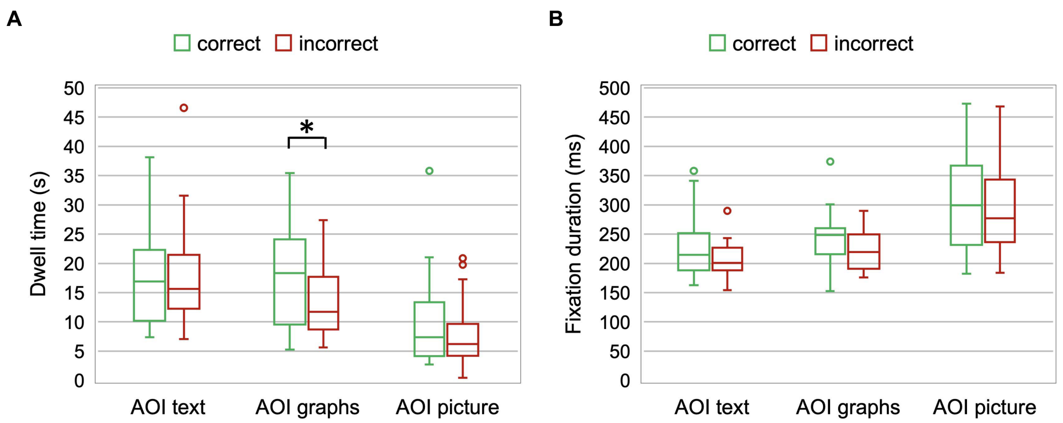


FIGURE 5
(A) Dwell time of students who correctly and incorrectly answered item Q2 for AOI text, AOI graphs, and AOI picture. **(B)** Average fixation duration of students who correctly and incorrectly answered item Q2 for AOI text, AOI graphs, and AOI picture.

information from the graphs. Also, students who possess prior conceptual knowledge related to the task (e.g., understanding the slope of the graph) may have an advantage in identifying and interpreting relevant information from the graphs, which could lead to more accurate problem-solving outcomes.

Overall, the answer to our RQ1 is that the main difference between students who correctly and incorrectly identify graphs from a verbal description lies in their examination of the offered graphs and extraction of relevant information. These results are consistent with the results of previous eye-tracking studies on students' understanding of graphs that have shown that understanding graphs and obtaining the required information is challenging for students, especially if it is very likely that they are not familiar with this type of tasks (Susac et al., 2018; Klein et al., 2019a,b).

To answer the research question RQ2 about the difference in the integration of information from multiple representations between students who correctly and incorrectly answer questions, we compared their number of transitions between different representations. The results show a trend of a higher number of transitions for students who solved the items correctly. However, the only statistically significant difference was found for transitions between AOI graphs and AOI picture in question Q2. This indicates that in that item it was crucial to connect the information from the picture and the graphs.

So, the answer to RQ2 is that students who were able to correctly answer the questions tended to have more transitions between representations, especially between those that were important for solving the items. This suggests that it is important for students to be able to link information from multiple representations in order to understand and answer PISA items that contain graphs.

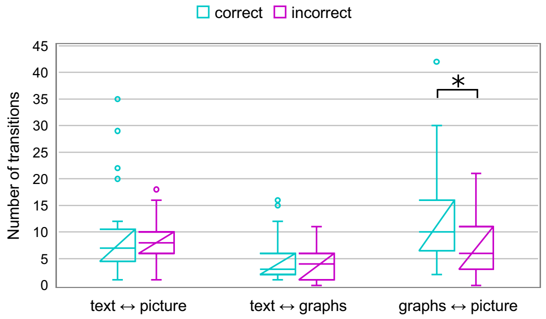


FIGURE 6
Number of transitions between AOI pairs (text ↔ picture, text ↔ graphs, and graphs ↔ picture) of students who correctly and incorrectly answered item Q2.

Sequence charts for Q1 and Q2 also illustrate that students who correctly solved these two PISA items spent more time analyzing the graphs than their peers who did not correctly solve the items. Furthermore, they show that students who are successful in solving items have a higher number of transitions between AOIs, which indicates a greater integration of information from different representations. Sequence charts also show high interindividual variability in measured data that could be a contributing factor to the failure to reach the statistical significance of some observed trends in eye-tracking measures for Q1 and Q2.

Our findings are consistent with those of a previous study by Mason et al. (2013), which identified three levels of integration of text and pictures and found that the greater the integrative processing of the illustrated text, the higher the learning performance. Ho et al. (2014) also found similar results in a study on how prior knowledge affects the processing of science texts containing graphs. They found that students with high prior knowledge had more regressions on the graphs, indicating that they were more able to integrate text and graphic information and effectively inspect scientific data. This ability to integrate multiple representations and examine data is essential for inquiry-based learning, and these results suggest that students with high prior knowledge may be better equipped to engage in this type of learning.

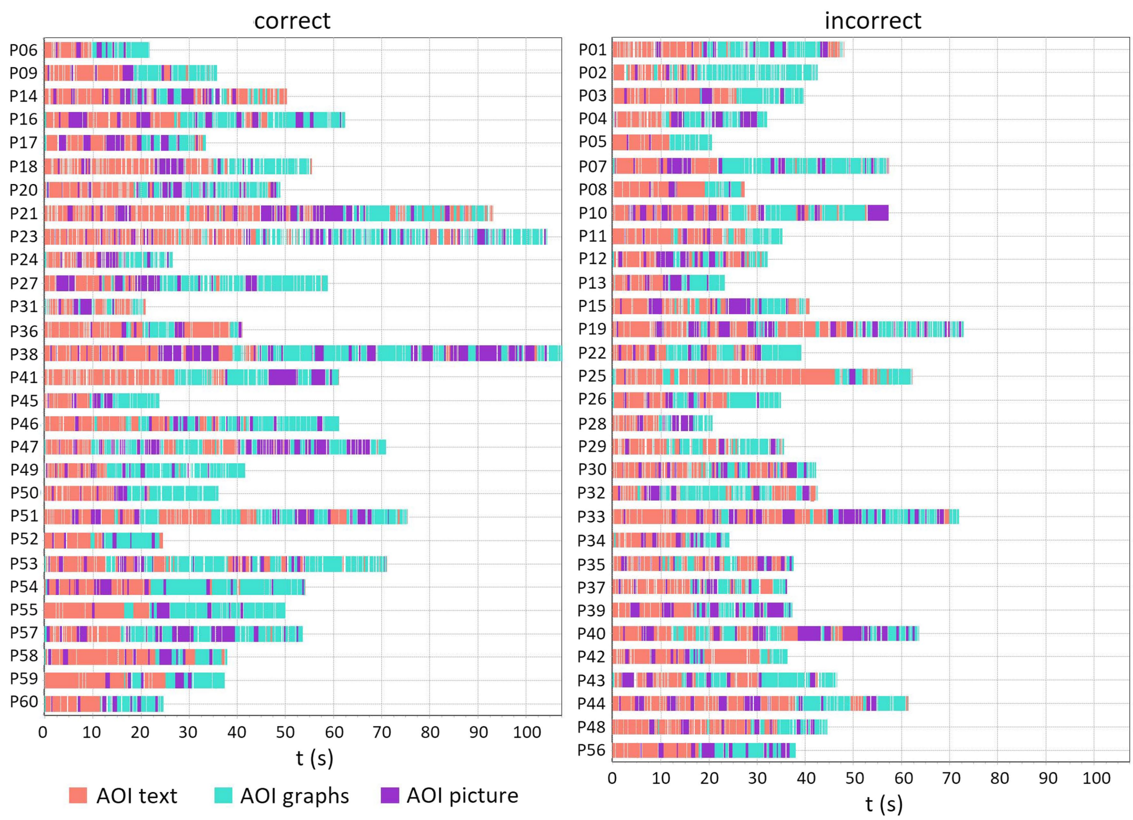


FIGURE 7
The sequence chart for PISA item Q2 shows the order of fixations for AOI text, AOI graphs, and AOI picture, separately for students who solved the item correctly and incorrectly. Fixations that fell within the boundaries of the AOIs are color-coded based on the color of the AOI. Each row represents a different participant.

The results of our study provided insight into students' visual attention during answering questions that required them to integrate information from different types of representations (such as text, graphs, and pictures). This information may be helpful for instructors in creating more effective teaching methods. By understanding how students are paying attention to and interacting with different representations, instructors can tailor their methods to better address the needs of their students, including the importance of carefully considering individual representations (especially graphs) and the need to integrate information from multiple sources.

When interpreting the results of this study, it is important to acknowledge that there are some limitations to consider. First, we analyzed students' eye movements while they were solving only two PISA items. The reason for this was that in other questions, the students had very high scores or the main cause of their difficulties was of a mathematical nature (e.g., problems with calculating percentages). In future studies, it would be needed to analyze the data for more different items to obtain more solid outcomes.

There are several reasons for such high scores of students on PISA items used in this study. University students were solving items intended for 15-year-olds. In addition, participants in our study did not represent the general student population, since we used convenience sampling. Only students who were ready to come to the university for research related to mathematics and physics participated in our study. In addition, our results showed that on some questions (e.g., M159 containing the graph of the speed of the racing car), students give correct answers because they easily eliminate other options, but it is not certain that they understand why the chosen option is correct. In future research, we plan to further investigate the observed problems with some PISA questions.

Furthermore, it would be desirable if we had an even larger sample of participants, although 60 participants is a fairly standard number of participants in eye-tracking studies. There are large differences between participants and the way they allocate their visual attention (Mason et al., 2013). Due to this great variability in the data, in order to obtain statistically significant differences in the results, it is necessary to have a larger number of participants. This is probably the cause of another limitation of this study, which is that some trends can be seen in the data, but they do not always reach statistical significance. Therefore, in future research, more different items with multiple representations should be used and a larger number of participants should be tested.

5. Conclusion

The results of the study suggest that students who are able to correctly solve PISA items that involve the integration of information from multiple representations (such as text, graphs, and pictures) tend to take longer to do so and make more transitions between these different representations than students who are not able to give correct answers. This indicates that these types of items require more time and effort to complete, particularly for students who are not familiar with this type of tasks. PISA items are not standard items that students encounter every day, so students do not have ready-made strategies for solving them. The study also found that the average fixation durations for different representations may vary, indicating that it is not always equally difficult to extract necessary information from different types of materials. These findings may be useful for instructors in developing more effective teaching methods that address the observed student behavior of needing to carefully consider individual representations and integrate information from multiple sources. By taking into account the importance of these factors, instructors may be able to better support their students in understanding and solving complex items.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AS, MPl, AB, KJ, and MPa contributed to the conception and design of the study. AS and MPa collected the data. AS analyzed the data and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learn. Instr.* 16, 183–198. doi: 10.1016/j.learninstruc.2006.03.001

Airy, J., and Linder, C. (2017). "Social semiotics in university physics education" in *Multiple Representations in Physics Education*. eds. D. Treagust, R. Duit and H. Fischer (Cham: Springer), 95–122.

Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., and Kuhn, J. (2020). Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: a Postreplication study. *Front. Psychol.* 11, 1–20. doi: 10.3389/fpsyg.2020.02090

Chen, S. C., and She, H. C. (2020). Effects of analogical learning approaches and presentation modalities on ninth graders' learning outcome and eye movements: a preliminary study. *J. Sci. Educ. Technol.* 29, 547–560. doi: 10.1007/s10956-020-09835-7

Chen, S. C., She, H. C., Chuang, M. H., Wu, J. Y., Tsai, J. L., and Jung, T. P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Comput. Educ.* 74, 61–72. doi: 10.1016/j.compedu.2013.12.012

Hahn, L., and Klein, P. (2022). Eye tracking in physics education research: a systematic literature review. *Phys. Rev. Phys. Educ. Res.* 18:013102. doi: 10.1103/PhysRevPhysEducRes.18.013102

Han, J., Chen, L., Fu, Z., Fritchman, J., and Bao, L. (2017). Eye-tracking of visual attention in web-based assessment using the force concept inventory eye-tracking of visual attention in web-based assessment using the force concept inventory. *Eur. J. Phys.* 38:045702. doi: 10.1088/1361-6404/aa6c49

Ho, H. N. J., Tsai, M.-J., Wang, C.-Y., and Tsai, C.-C. (2014). Prior knowledge and online inquiry-based science reading: EVIDENCE from eye tracking. *Int. J. Sci. Math. Educ.* 12, 525–554. doi: 10.1007/s10763-013-9489-6

Hoyer, C., and Girwidz, R. (2020). Animation and interactivity in computer-based physics experiments to support the documentation of measured vector quantities in diagrams: an eye tracking study. *Phys. Rev. Phys. Educ. Res.* 16:020124. doi: 10.1103/PhysRevPhysEducRes.16.020124

Hu, Y., Wu, B., and Gu, X. (2017). An eye tracking study of high-and low-performing students in solving interactive and analytical problems. *Educ. Technol. Soc.* 20, 300–311.

Ibrahim, B., and Ding, L. (2021). Sequential and simultaneous synthesis problem solving: a comparison of students' gaze transitions. *Phys. Rev. Phys. Educ. Res.* 17:10126. doi: 10.1103/PhysRevPhysEducRes.17.010126

Kekule, M. (2014). Students' approaches when dealing with kinematics graphs explored by eye-tracking research method. *Eur. J. Sci. Math. Educ.* 2, 108–117. doi: 10.30935/scimath/9632

Kekule, M., and Viiri, J. (2018). Students' approaches to solving R-FCI tasks observed by eye-tracking method. *Sci. Educ.* 9, 117–130. doi: 10.14712/18047106.1010

Klein, P., Küchemann, S., Brückner, S., Zlatkin-Troitschanskaia, O., and Kuhn, J. (2019a). Student understanding of graph slope and area under a curve: a replication study comparing first-year physics and economics students. *Phys. Rev. Phys. Educ. Res.* 15:20116. doi: 10.1103/PhysRevPhysEducRes.15.020116

Klein, P., Lichtenberger, A., Kuchemann, S., Becker, S., Kekule, M., Viiri, J., et al. (2019b). Visual attention while solving the test of understanding graphs in kinematics: an eye-tracking analysis. *Eur. J. Phys.* 41:025701. doi: 10.1088/1361-6404/ab5f51

Klein, P., Viiri, J., and Kuhn, J. (2019c). Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving. *Phys. Rev. Phys. Educ. Res.* 15:10126. doi: 10.1103/PhysRevPhysEducRes.15.010126

Klein, P., Viiri, J., Mozaffari, S., Dengel, A., and Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: how do students look at vector field plots? *Phys. Rev. Phys. Educ. Res.* 14:10116. doi: 10.1103/PhysRevPhysEducRes.14.010116

Kozhevnikov, M., Motes, M. A., and Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cogn. Sci.* 31, 549–579. doi: 10.1080/15326900701399897

Krstić, K., Šoškić, A., Ković, V., and Holmqvist, K. (2018). All good readers are the same, but every low-skilled reader is different: an eye-tracking study using PISA data. *Eur. J. Psychol. Educ.* 33, 521–541. doi: 10.1007/s10212-018-0382-0

Küchemann, S., Klein, P., Fouckhardt, H., Gröber, S., and Kuhn, J. (2020). Students' understanding of non-inertial frames of reference. *Phys. Rev. Phys. Educ. Res.* 16:10112. doi: 10.1103/PhysRevPhysEducRes.16.010112

Langendorf, R., Schneider, S., and Klein, P. (2022). Extracting information from the Hertzsprung-Russell diagram: an eye-tracking study. *Phys. Rev. Phys. Educ. Res.* 18:020121. doi: 10.1103/PhysRevPhysEducRes.18.020121

Lundgren, E. (2022). Latent program modeling: Inferring latent problem-solving strategies from a PISA problem-solving task. *J. Educ. Data Mining* 14, 46–80. doi: 10.5281/zenodo.6686443

Madsen, A. M., Larson, A. M., Loschky, L. C., and Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Phys. Rev. Spec. Top. Phys Educ. Res.* 8:010122. doi: 10.1103/PhysRevSTPER.8.010122

Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., and Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Phys. Rev. Spec. Top. Phys Educ. Res.* 9:020104. doi: 10.1103/PhysRevSTPER.9.020104

Mason, L., Tornatora, M. C., and Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Comput. Educ.* 60, 95–109. doi: 10.1016/j.compedu.2012.07.011

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educ. Psychol.* 32, 1–19. doi: 10.1207/s15326985ep3201_1

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., and Plass, J. L. (2014). Learning from multiple representations: an examination of fixation patterns in a science simulation. *Comput. Hum. Behav.* 35, 234–242. doi: 10.1016/j.chb.2014.02.040

OECD (2006a). PISA released items-mathematics.

OECD (2006b). PISA released items-science.

OECD (2007). PISA 2006: science competencies for tomorrow's world. Available at: www.sourceoecd.org/9789264040007

OECD (2018). *The Future of Education and Skills*. Paris: OECD Publishing

Ott, N., Brünken, R., Vogel, M., and Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learn. Instr.* 58, 88–105. doi: 10.1016/j.learninstruc.2018.04.010

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart, and Winston

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educ. Psychol. Rev.* 29, 717–761. doi: 10.1007/s10648-016-9365-3

Schnotz, W., and Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learn. Instr.* 13, 141–156. doi: 10.1016/S0959-4752(02)00017-8

Skrabankova, J., Popelka, S., and Beitlova, M. (2020). Students' ability to work with graphs in physics studies related to three typical student groups. *J. Balt. Sci. Educ.* 19, 298–316. doi: 10.33225/jbse/20.19.298

Stieff, M., Hegarty, M., and Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cogn. Instr.* 29, 123–145. doi: 10.1080/07370008.2010.507318

Susac, A., Bubic, A., Kazotti, E., Planinic, M., and Palmovic, M. (2018). Student understanding of graph slope and area under a graph: a comparison of physics and nonphysics students. *Phys. Rev. Phys. Educ. Res.* 14:020109. doi: 10.1103/PhysRevPhysEducRes.14.020109

Susac, A., Bubic, A., Martinjak, P., Planinic, M., and Palmovic, M. (2017). Graphical representations of data improve student understanding of measurement and uncertainty: an eye-tracking study. *Phys. Rev. Phys. Educ. Res.* 13:020125. doi: 10.1103/PhysRevPhysEducRes.13.020125

Susac, A., Bubic, A., Planinic, M., Movre, M., and Palmovic, M. (2019). Role of diagrams in problem solving: an evaluation of eye-tracking parameters as a measure of visual attention. *Phys. Rev. Phys. Educ. Res.* 15:13101. doi: 10.1103/PhysRevPhysEducRes.15.013101

Susac, A., Planinic, M., Bubic, A., Ivanjek, L., and Palmovic, M. (2020). Student recognition of interference and diffraction patterns: an eye-tracking study. *Phys. Rev. Phys. Educ. Res.* 16:20133. doi: 10.1103/PhysRevPhysEducRes.16.020133

Susac, A., Planinic, M., Bubic, A., Jelicic, K., Ivanjek, L., Matejak Cvenic, K., et al. (2021). Effect of students' investigative experiments on students' recognition of interference and diffraction patterns: an eye-tracking study. *Phys. Rev. Phys. Educ. Res.* 17:10110. doi: 10.1103/physrevphyseducres.17.010110

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.* 12, 257–285. doi: 10.1016/0364-0213(88)90023-7

Thomaneck, A., Vollstedt, M., and Schindler, M. (2022). What can eye movements tell about students' interpretations of contextual graphs? A methodological study on the use of the eye-mind hypothesis in the domain of functions. *Front. Educ.* 7:1003740. doi: 10.3389/feduc.2022.1003740

Tóthová, M., and Rusek, M. (2022). Do you just have to know that? Novice and experts' procedure when solving science problem tasks. *Front Educ* 7:1051098. doi: 10.3389/feduc.2022.1051098

Van Gog, T., Paas, F., and van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Appl. Cogn. Psychol.* 19, 205–221. doi: 10.1002/acp.1112

Wu, C. J., and Liu, C. Y. (2021). Eye-movement study of high-and low-prior-knowledge students' scientific argumentations with multiple representations. *Phys. Rev. Phys. Educ. Res.* 17:10125. doi: 10.1103/PhysRevPhysEducRes.17.010125

frontiers | Frontiers in Education

TYPE Review
PUBLISHED 03 November 2023
DOI 10.3389/feduc.2023.1170348

OPEN ACCESS

EDITED BY
Pascal Klein,
University of Göttingen, Germany

REVIEWED BY
Habiddin Habiddin,
State University of Malang, Indonesia
Shuai Liu,
Hunan Normal University, China

*CORRESPONDENCE
Yakhoub Ndiaye
✉ yakhoub_ndiaye@sutd.edu.sg

RECEIVED 20 February 2023
ACCEPTED 11 October 2023
PUBLISHED 03 November 2023

CITATION
Ndiaye Y, Lim KH and Blessing L (2023) Eye tracking and artificial intelligence for competency assessment in engineering education: a review.
Front. Educ. 8:1170348.
doi: 10.3389/feduc.2023.1170348

Eye tracking and artificial intelligence for competency assessment in engineering education: a review

Yakhoub Ndiaye[1]*, Kwan Hui Lim[2] and Lucienne Blessing[3]

[1]Department of SGP AI, Singapore University of Technology and Design, Singapore, Singapore, [2]Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore, Singapore, [3]Pillar of Engineering Product Development, Singapore University of Technology and Design, Singapore, Singapore

In recent years, eye-tracking (ET) methods have gained an increasing interest in STEM education research. When applied to engineering education, ET is particularly relevant for understanding some aspects of student behavior, especially student competency, and its assessment. However, from the instructor's perspective, little is known about how ET can be used to provide new insights into, and ease the process of, instructor assessment. Traditionally, engineering education is assessed through time-consuming and labor-extensive screening of their materials and learning outcomes. With regard to this, and coupled with, for instance, the subjective open-ended dimensions of engineering design, assessing competency has shown some limitations. To address such issues, alternative technologies such as artificial intelligence (AI), which has the potential to massively predict and repeat instructors' tasks with higher accuracy, have been suggested. To date, little is known about the effects of combining AI and ET (AIET) techniques to gain new insights into the instructor's perspective. We conducted a Review of engineering education over the last decade (2013–2022) to study the latest research focusing on this combination to improve engineering assessment. The Review was conducted in four databases (Web of Science, IEEE Xplore, EBSCOhost, and Google Scholar) and included specific terms associated with the topic of AIET in engineering education. The research identified two types of AIET applications that mostly focus on student learning: (1) eye-tracking devices that rely on AI to enhance the gaze-tracking process (improvement of technology), and (2) the use of AI to analyze, predict, and assess eye-tracking analytics (application of technology). We ended the Review by discussing future perspectives and potential contributions to the assessment of engineering learning.

KEYWORDS

eye tracking, artificial intelligence, competency, assessment, engineering education

1. Introduction

Eye tracking has been integrated into many applications, such as human-computer interaction, marketing, medicine, and engineering (e.g., assistive driving, software, and user interfaces). Recent studies revealed that eye tracking (ET) and artificial intelligence (AI), including machine (ML) and deep learning (DL), have been combined to assess human behavior (e.g., Tien et al., 2014). However, although extensive studies have focused on the application of AI techniques to eye-tracking data in some STEM disciplines, little is known about how this could be used in engineering design educational settings to facilitate instructors' assessment of the design learning of their students, especially design competency.

1.1. Competency assessment in engineering education

1.1.1. Competency-based engineering education

The development of student competencies has become a central issue in complex fields, such as engineering education. With regard to competencies, various terminologies are used to describe a learner expertise in a situation and their ability to solve complex engineering problems; for instance, competence (pl. competences), competency (pl. competencies), capability, and so on, are generally used. The debate about terminology is still ongoing. In this paper, we refer to both "competence," i.e., the general term, and "competency," i.e., the components of a competence as holistic constructs, with the focus on "competency" as the ability to integrate knowledge, skills, and attitudes (KSAs; Le Deist and Winterton, 2005) and their underlying constituents (cognitive, conative, affective, motivational, volitional, social, etc.; e.g., Shavelson, 2013; Blömeke et al., 2015) simultaneously (van Merriënboer and Kirschner, 2017). From an instructional design perspective, learning, which is also the acquisition of skills and competencies, has integrative goals in which KSAs are developed concurrently to acquire complex skills and professional competencies (Frerejean et al., 2019). This approach is interesting and may help avoid core issues in instructional engineering design, such as compartmentalization, which involves the teaching of KSAs separately, hindering competency acquisition and transfer in complex engineering learning. Therefore, as suggested by Spencer (1997), competency assessment (we discuss this further in the next section) determines the extent to which a learner has competencies. Competency is assumed to be multidimensional (Blömeke et al., 2015) and discipline-specific (Zlatkin-Troitschanskaia and Pant, 2016). Competencies can be learned through training and practice. Siddique et al. (2012) noted two levels of competencies in professional fields: (1) field-specific task competencies, and (2) meta-competencies as generalized skill sets. Le Deist and Winterton (2005) argued that a multi-dimensional framework of "competence" necessarily involved conceptual (cognitive and meta-competence) and operational (functional and social competence, including attitude and behavior) competencies. They assumed competence is composed of four dimensions of competencies: cognitive dimension (knowledge), functional dimension (skills), social (behavior and attitudes), and meta-competence (Le Deist and Winterton, 2005). Engineers argue that these dimensions also apply to engineering education. With the emerging complexity involved in designing engineering systems, tackling complexity is a new requirement. As such, Hadgraft and Kolmos (2020) proposed that three basic competencies should be incorporated into engineering courses: complexity, system thinking, and interdisciplinarity. Therefore, we argue that competency and competency assessment should be described by a more holistic framework that is appropriate to learning and instruction in complex engineering education.

1.1.2. Challenges of assessing student competencies in engineering education

Instructors' assessment of students' engineering competencies is a critical topic that has been addressed for decades in the engineering education literature. Despite this, assessment of engineering learning suffers from several issues, such as a lack of consistency. It is still highly subjective, labor-intensive, and time-consuming. The COVID-19 pandemic has exacerbated these issues as many engineering instructions shifted from face-to-face to online or remote instructions using online platforms, thus increasing teacher workloads, cognitive loads, etc. More critically, engineering assessment suffers from an integrative approach to engineering competencies and competency assessment even with the use of advanced techniques, such as AI and other computing technologies (e.g., Khan et al., 2023). Most technologies used to assess engineering student competencies usually focus on some aspects of an engineering competence and not on a systemic holistic approach to competency.

1.2. Eye-tracking technologies: a brief history in scientific research

Several papers have reviewed the history of eye-tracking research (e.g., Wade and Tatler, 2005; Płużyczka, 2018). Płużyczka (2018) identified three developmental phases in the first 100 years of eye tracking as a research approach: the first phase of eye-tracking research dates back to the late 1870s with Javal's studies on understanding and assessing the reading process. At that time, the eye-tracking approach was optical-mechanical and invasive. The second era of eye-tracking research originated with film recordings in the 1920s. The third phase started in the mid-1970s and refers to two main phenomena related to the development of psychology (the establishment of a theoretical and methodological basis for cognitive psychology) and technology (the use of computer, television, and electronic techniques to detect and locate the eye). Motivated by the rapid development of eye-tracking and computer processing technologies, Płużyczka (2018) also suggested that another phase led to contemporary eye-tracking research that took place in the 1990s.

Eye tracking permits the assessment of an individual's visual attention, yielding a rich source of information on where, when, how long, and in which sequence certain information in space or about space is looked at (Kiefer et al., 2017). Different eye-tracking techniques have been referenced. For instance, Duchowski and Duchowski (2007) identified four categories of eye movement measurement methodologies: electro-oculography (EOG), scleral contact lens/search coil, photo-oculography (POG) or video-oculography (VOG), and video-based combined pupil and corneal reflection (p. 51). Li et al. (2021) also provided a similar overview. Among other techniques, they cited the earliest manual observations followed by new techniques, such as electrooculography, video and photographic, corneal reflection, and micro-electromechanical systems, and those based on machine and deep learning. For each method, they examined the benefits and limitations. They argued that CNN-based approaches offer better recognition performance and robustness; however, they require large amounts of data, complex parameter adjustments, and an understanding of black box characteristics, and involve high costs.

1.3. AI and ET to assess engineering

Artificial intelligence and computer vision (CV) have advanced significantly and rapidly over the past decade due to highly effective deep learning models, such as the CNN variants (Szegedy et al., 2015; He et al., 2016; Huang et al., 2017) and vision transformers (Dosovitskiy et al., 2020), and the availability of large high-quality datasets and powerful GPUs for training such large models. As a result of these advances in AI and CV, eye-tracking technology has reached a level of reliability sufficient for wider adoption, such as for evaluating student attention via their eye-gaze on the study materials taught. More specifically, this application has the benefit of being able to measure multiple spectrums of student attention. For example, such technology can measure whether the student is focusing more generally on the class or specifically on certain parts of the lecture material. Adding on the dimension of time, one can also measure the amount of time students spend on different parts of the course content and when their attention starts to drift.

In terms of the instructor-side, the integration of eye tracking and AI has various benefits for the assessment of engineering design education. Similar to the application of eye tracking and AI with students, these technologies can generally be used to measure which part of a student assignment an instructor focuses more on and the amount of time they spend on different parts of an assignment. In addition, we see the following potential cases for the use of eye tracking and AI:

- Studying the effectiveness of assessment criterions. Alongside a marking rubric, eye tracking and AI can be used to find a correlation between different assessment criteria and specific parts of a submitted assignment. For example, we can compare the criteria in a marking rubric that an instructor is looking at and the corresponding parts of an assignment they look at next. Pairs of these marking criteria and assignment segments can then be used for correlation studies.
- Streamlining instructor assessment workload. With explainable AI (XAI) techniques, a system can highlight portions of the student assignments that an instructor should focus on based on the different criteria. Such a model can be trained on past data of instructor assessment and student assignments, alongside the captured eye-tracking data. This model can then be transferred and fine-tuned to other assignments.
- Detecting discrepancies between instructor assessments. Different instructors may have varying standards or interpretations of engineering assessments, e.g., between newer and more experience instructors. Eye tracking and AI can be used to determine whether there are any differences between instructors in terms of the parts of the student assignments they focus on, how much time they spend on each portion, and, most importantly, any significant differences in the assigned grades for each criteria.

2. Purpose and research questions

This study aims to understand research trends in the use of AIET to assess engineering student competencies. The overall research questions (RQs) are as follows:

- RQ1: What are the current research trends (or categories) in AI- and ET-based competency assessment in engineering education over the last decade?
- RQ2: What are the most salient competency dimensions and labels to which we attribute studies related to assessing engineering education?

3. Methods

3.1. Data collection

We reviewed the literature and collected papers from the following four databases: Web of Science (WoS), IEEE Xplore (IX), Academic Search Complete (ASC), and Computers and Applied Sciences Complete (CASC) hosted by EBSCO and Google Scholar (GS). The Review was conducted with research published in the last decade, i.e., from 2013 to 2022. Focusing on title, abstracts, and keywords, we used a general equation including terms used in the topic of eye tracking and artificial intelligence in engineering education research, such as Title-Abs-Key[("eye-track*" OR "eye-gaze" OR "eye movement") AND ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("assess* OR evaluat* OR measur* OR test* OR screen*) AND ("competenc*" OR "skills" OR "knowledge" OR "attitudes") AND ("engineering design" OR "engineering education")]. The review process, which comprised three steps, namely identification, screening, and eligibility, is summarized below and in the flow diagram (Figure 1):

1. Identification: an initial record of $N = 89$ studies were identified by searching the databases: EBSCOhost (19 studies), WoS (24 studies), IX (26 studies), and GS (20 studies).
2. Screening: after duplicates were removed, records were screened based on the relevance of titles and abstracts.
3. Eligibility: peer-reviewed studies written in English and related to engineering education, competency assessment, and higher education were selected.
4. Finally, $N = 76$ studies were retained in this Review.

All references collected from the databases were imported into Rayyan, an intelligent platform for systematic review, to help in the review process. Data were then manually categorized (according to labels that fit in the dimensional aspects of a student's competency as defined earlier) and exported in an editable format containing three variables: title, abstract, dimension, and corresponding labels. In addition, the generated format contained the following criteria: relevance to assessing EE (yes/no), higher education (yes/no), tested with instructors or students (yes/no), methodology used (type of assessment), competency dimensions (cognitive, functional, social, and meta), contributions, and limitations (Figure 2).

3.2. Data analysis

Qualitative and quantitative methods were used to analyze the data. Following the collection, we first performed a qualitative analysis (i.e., thematic analysis), manually categorizing and labeling the focus of each paper according to the competency dimensions. This helped

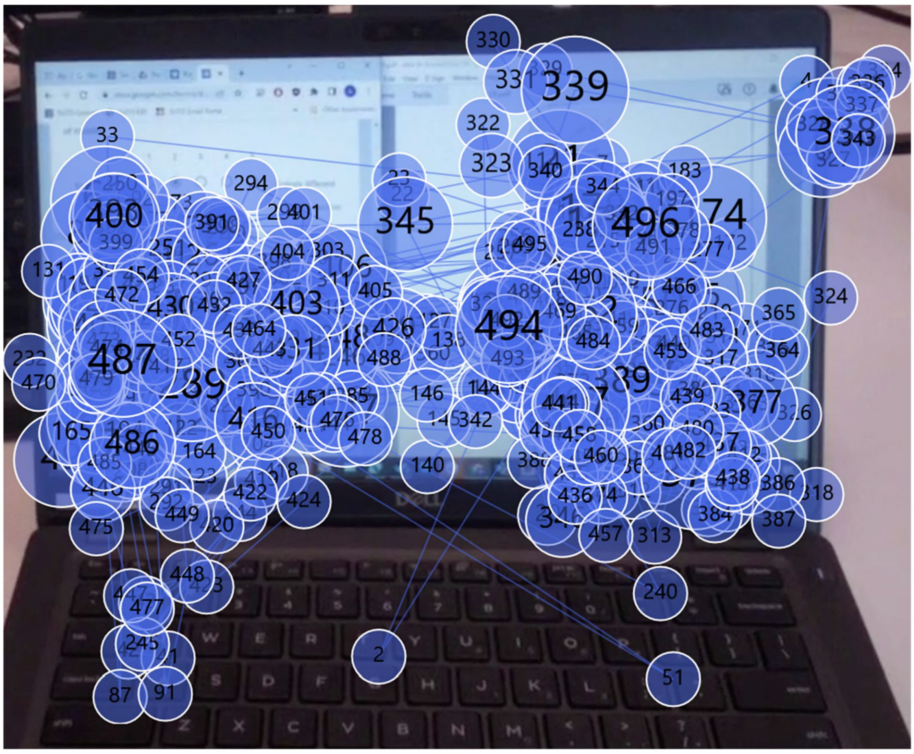


FIGURE 1
Instructor on-screen assessment of student design creativity. This pilot experiment (one person involved) used Tobii glasses 2 (issues related to the use of this specific tool are not discussed here). The instructor followed a rubric (left side of the screen) comprising a set of criteria to assess creativity in students' design solutions (right side of the computer screen). The results of this study are not reported in this paper.

to identify the types of AIET. Based on this corpus, we then furthered the Review with a lexicometry analysis with IRaMuTEQ 0.7 alpha 2 (Ratinaud, 2009) and RStudio 2021.09.1 + 372 for macOS. IRaMuTEQ is an R interface for multidimensional analysis of texts and questionnaires. It offers different types of analysis, such as lexicometry, statistical methods (specificity calculation, factor analysis, or classification), textual data visualization (usually called word cloud), or term network analysis (called similarity analysis).

We conducted a clustering based on the Reinert's method (Reinert, 1990). This method includes a hierarchical classification, profiles, and correspondence analyses. To obtain the co-occurrence graphs, we then conducted a similarities analysis that used the graph theory also called network analysis to analyze trends within the reported data. Finally, we also used thematic analysis to organize the reported data into categories for the assessment types, titles of clusters, and types of AIET.

4. Results

4.1. Categories of artificial intelligence and eye tracking

Our first research question attempts to explore current research trends in AIET-based assessment in engineering education. Typically, and based on the manual thematic analysis, two relatively dependent types of AIET research categories can be identified with regard to assessment: (1) eye-tracking devices that use AI and sub-domains to improve the process of tracking (improving the technology), and (2) the use of AI to analyze and predict the eye-tracking data analytics related to student learning (the application of the technology). The first typology generally consists of combining AI and sub-domains, such as machine learning (ML) and deep learning (DL) with ET. As opposed to traditional tracking approaches that often estimate the location of visual cues, researchers developing this orientation attempt to improve the tracking process; for instance, favoring detection over tracking. Reported results from this approach detail the performance and accuracy of detection. This is receiving increasing attention. Conversely, although the second typology also utilizes AI to predict and detect behaviors, it mainly focuses on assessing and providing insights into learner behaviors afterwards based on recorded eye-tracking data. Collected data can be reinjected into the learning system afterwards to support the learners and/or educators.

Usually there are more practical applications to educational assessment. With regard to these typologies, a multimodal approach integrating ET and several signals, such as EEG (e.g., Wu et al., 2021), fNIRS (Shi et al., 2020), and skin conductance (e.g., Muldner and Burleson, 2015), is also referenced (Table 1).

4.2. Engineering competencies and dimensions

We view competency as the integration of student skills, knowledge, and attitudes and their underlying constituents

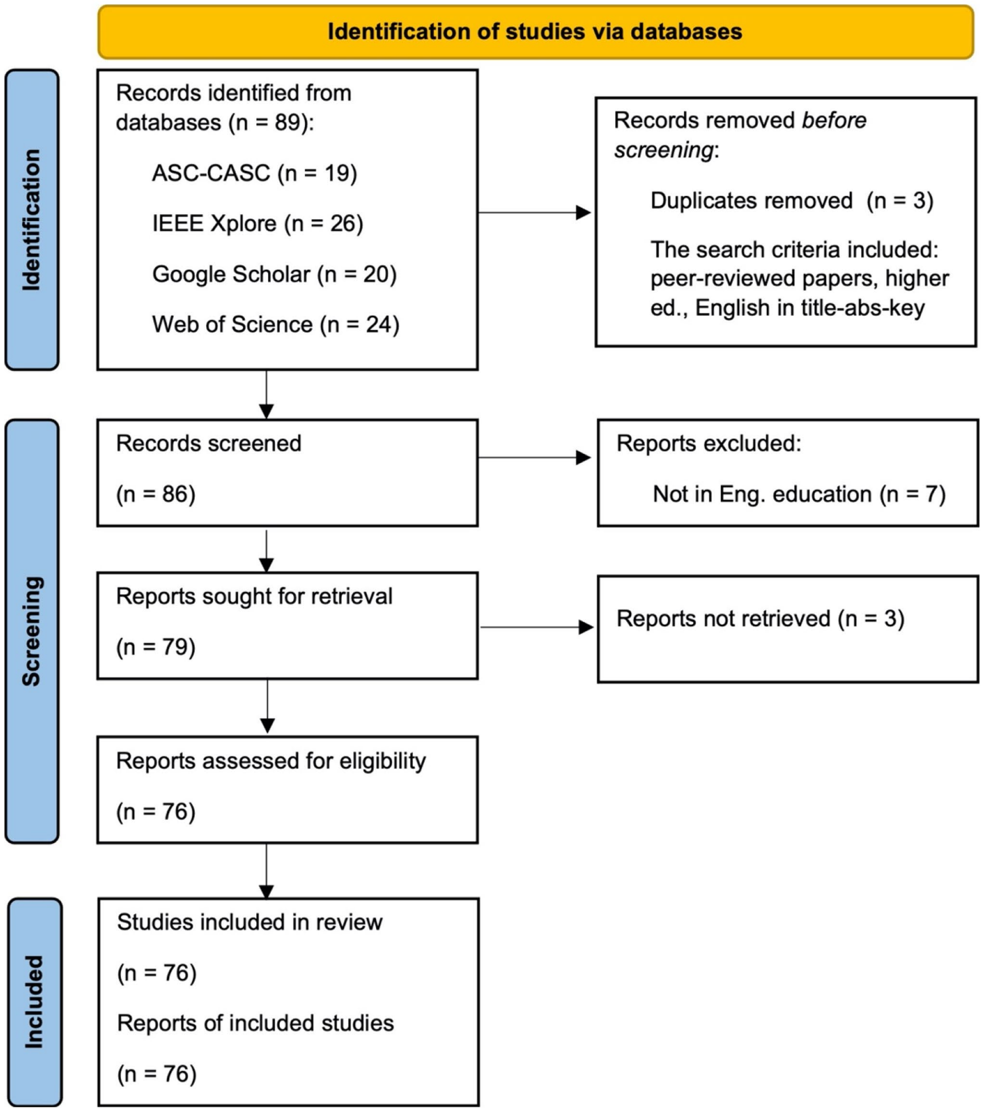


FIGURE 2
Flow diagram.

simultaneously, hence highlighting different learning dimensions as defined earlier. There is no meaningful skill acquisition without suitable connections to these defined dimensions. Consequently, the competency acquisition is analyzed in terms of these dimensions, namely cognitive, functional, social, and meta. Among these dimensions, the assessment of the cognitive dimension of engineering student expertise seems to be the primary focus of AIET applications (*cf.* Table 2).

Moreover, our results showed an overview of learners' mental state assessments, including cognitive, affective, and social levels of learners' competencies. Although the visual and cognitive competency dimension was a particular focus, studies are lacking when it comes to students' design expertise and its assessment by instructors. The reported studies examined the issues of addressing an aspect of student competency; however, they still lack focus on a holistic approach of competency assessment with competency being the integration of KSAs. Additionally, we manually analyzed references to highlight the types of assessment included in studies (see Table 3). The lexicometry analysis ran a hierarchical top-down classification that helped to identify four classes (or clusters) within the reported data (see Table 4). Class 1 (28.1% of the data), which we named "Eye-tracking method," grouped terms related to gaze, achieve, feature, and eye, which were used to track and assess visual patterns. This class is correlated with Class 2 (13.9%) comprising the "AI functional approach" used to track and assess learners' mental states and behaviors through the eye-tracking analytics. Such behaviors are described in Class 3 (22.1%), which we named "Mental state and behavioral assessment." This class included terminology associated with the assessed aspect/behavior, such as cognitive, perceptual, awareness, stress, mental, and competency. Finally, we identify Class 4 (35.9%), which addressed "Instructional approach and student learning" as it included terms such as "student," "learning," and "team."

In addition to this classification, we ran a correspondence analysis (CA) that showed the visual relationships of the identified clusters (*cf.* Figure 3). We analyzed the CA based on the first two

TABLE 1 Typologies of AI and ET.

Dimensions	Focuses	Examples	References
Using AI to improve the tracking technology	Tracking the reading progression: line detection vs. line tracking	Although traditional eye trackers provide an estimation of the eye-gaze points and their location every few milliseconds (not sufficient to quantify reading progression), this approach uses a Kalman filter and hidden Markov model to detect read lines accurately. The estimated eye-tracking point improved line detection accuracy by 27.1% relative to line tracking.	Bottos and Balasingam (2020)
Using AI to analyze and predict the eye-tracking data	Prediction of the difficulty level of spatial visualization problems	The use of machine learning to study (1) the differences in eye movement between different difficulty levels of the problem and (2) the possibility of predicting the difficulty level from eye-tracking data. The model generated an average accuracy of 87.60% for tracking data seen by the classifier, and 72.87% for unseen data.	Li et al. (2020)

TABLE 2 Cognitive, functional, social and meta aspects of competency assessment.

Dimension	Categories (% Freq.)	References
Cognitive (64.7%)	Spatial visualization, design behaviors (3.2%)	Muldner and Burleson (2015), Dogan et al. (2018), Li et al. (2020), and Mehta et al. (2020)
	Measuring cognitive loads (9.7%)	Bozkir et al. (2019) and Amadori et al. (2021)
	Attention (25.8%), concentration, and engagement (1.6%)	Meza et al. (2017), Guo and Barmaki (2020), Bharadva (2021), Chakraborty et al. (2021), Su et al. (2021), Khosravi et al. (2022), Renawi et al. (2022), and Singh and Modi (2022)
	Cognitive vigilance and awareness (16.1%)	Farha et al. (2021) and Lili et al. (2021)
	Comprehension, retention (6.5%), and perception of behavior (3.2%)	Das and Hasan (2014) and Hijazi et al. (2021)
Functional (17.7%)	Reading skills, speaking proficiency (1.6%)	Bottos and Balasingam (2020) and Tamim et al. (2021)
	Classification of learning (1.6%)	Pritalia et al. (2020)
	Recognition of creativity skills (6.5%)	Muldner and Burleson (2015)
	Navigation (3.2%), traceability (1.6%), and decision making (1.6%)	Ahrens (2020) and Lili et al. (2021)
Social (14.5%)	Affective and emotion recognition (9.7%)	Aracena et al. (2015) and Meza et al. (2017)
	Interpersonal skills (e.g., teamwork, communication; 4.8%)	Amri et al. (2017), Chen (2021), and Lili et al. (2021)
Meta (3.2%)	Intention to cheat (1.6%)	Singh and Das (2022)

TABLE 3 Types of assessment.

Types	Studies	Example of tasks
Formative assessment	Guo and Barmaki (2020), Su et al. (2021) and Tamim et al. (2021)	Automatic assessment of team performance during collaborative tasks
Summative assessment	Bottos and Balasingam (2020), Bozkir et al. (2019), Ahrens (2020), and Hijazi et al. (2021)	iReview, an intelligent tool used to evaluate code reviews
Self-assessment	Khosravi et al. (2022)	Learners can use the eye tracker for attention guidance
Peer-assessment	Chen (2021)	TeamDNA, used to measure the communication aspect of teamwork. It provides objective and non-interruptive measurements, observer-based measures with team process-based analyses, and sensor-based measures with non-intrusive measurements

factors, which were quite representative of the data samples (factor 1, 40.81%; factor 2, 32.19%). Results highlighted that Clusters 1 and 2 were well correlated, suggesting the relevance of the association of eye tracking and AI. However, these two clusters were in opposition, i.e., negatively correlated with Cluster 3 about mental states and behavior assessment on axis 2 (vertical), and with Cluster 4 about instructional approaches on axis 1. From these results, we identified the most well represented words of each cluster (see Table 4; "gaze" in cluster 1: $\chi^2 = 33.78$, $p < 0.0001$; "network" in cluster 2: $\chi^2 = 258.56$, $p < 0.0001$; "cognitive" in cluster 3: $\chi^2 = 51.41$, $p < 0.0001$; and "student" in cluster 4: $\chi^2 = 88.44$, $p < 0.0001$). This analysis confirmed the three clusters, namely the classes described above.

To obtain the co-occurrence graphs, we then performed a similarities analysis that used the graph theory also called network analysis to analyze trends in the literature. This analysis displayed the overall connection and grouping of terms used in the reported papers based on the co-occurrence scores of words (*cf.* Figures 4–6).

TABLE 4 Significancy table (terms per class).

Clusters: name	Most significant terms per cluster*	Chi-square χ^2 (p-value)	Term sources (or correlated with)
Cluster 1: "Eye-tracking approach"	Gaze	33.78 (<0.0001)	-
	Achieve	33.35 (<0.0001)	
28.07%	Feature	28.86 (<0.0001)	
	Eye	23.52 (<0.0001)	
Cluster 2: "AI functional approach"	Network	258.56 (<0.0001)	Keywords (χ^2 = 4.84; p = 0.02778)
	Neural	224.43 (<0.0001)	
13.94%	Convolutional	95.26 (<0.0001)	
	Emotion	64.55 (<0.0001)	
Cluster 3: "Mental state and behavior assessment"	Cognitive	51.41 (<0.0001)	-
	Drive	33.72 (<0.0001)	
22.12%	Perceptual	28.59 (<0.0001)	
	Awareness	26.59 (<0.0001)	
Cluster 4: "Instructional approach and student learning"	Student	88.44 (<0.0001)	Abstract (NS***; p = 0.10722)
	Learn	42.66 (<0.0001)	
35.85%	Team	35.4 (<0.0001)	
	Online	29.67 (<0.0001)	

*Due to the limitation of table dimension, only the first four significant terms are provided. **The terms "assessment" and "assess" were situated in the cluster 3 list, with χ^2 = 16.17 (<0.0001) and 15.96 (<0.0001). ***NS, not significant.

Whereas Figure 4 highlights grouping words from the reported literature (title, abstract, and keyword), Figure 6 shows the relationships between those three variables and our defined dimensions (cognitive, functional, social, and meta) and the labels defined in Table 2.

Regarding competency assessment in engineering education, we particularly focused on the feature "assess*" and analyzed (1) trends (Figure 7), (2) keyword-in-context (Table 5). With regard to research trends, Figure 7 (top), which depicts the absolute occurrence of the feature "assess*" across years, shows a clear trend of increasing interest in assessment with technologies, such as AI and ET, whereas Figure 7 (bottom) outlines the relative occurrence of "assess*" in comparison with all other features. Examples of citations in abstracts mentioning the purpose of analysis are provided in Table 4.

4.3. Application scenarios and model accuracy

Artificial intelligence and eye tracking has been applied in several engineering contexts with different focuses. A summary is provided in Table 6. We noticed applications in the classroom but also in lab practice, simulation training, and industry. However, despite the relevance, research is lacking on how models can help assess competency in a broader way involving the dimensions discussed earlier.

To understand the relevance of these AIET applications, we identified studies on the accuracy of developed models that integrate AI and ET to support engineering assessment broadly speaking, i.e., of and for/as learning with regard to the engineering literature (see Table 7). A relatively good average accuracy of 79.76% was found with an estimation range from 12% to 99.43%.

5. Discussion

This paper reviews the engineering literature to identify research focusing on AI and ET to support the assessment of competency in engineering education. Our study revealed that combining eye tracking and AI to assess engineering student competencies is receiving increasing attention. The association seems to be well supported, especially with the development of advanced technologies, such as AI. The Review highlights the main types of AIET, which are discussed below.

5.1. Two types of AIET in engineering competency assessment

Overall, two types of AIET focuses were reported: (1) an eye-tracking device that uses AI to improve the process of visual tracking itself, and (2) the use of AI to analyze, predict, and assess the eye tracking analytics.

5.1.1. AIET to improve the process of visual tracking

Most studies reported in this Review use this first approach to improve current eye-tracking technologies. For instance, in recent years, the prediction of eye movement scanpath can be divided into two categories: prediction models that hand-design features and powerful mathematical knowledge, and methods that intuitively obtain the sequence of eye fixes from the bottom-up salinity map and other useful indications (Han et al., 2021). With the advances of machine and deep learning, the study of computational eye-movement models has been mainly based on neural network learning models (e.g., Wang et al., 2021). For instance, in the

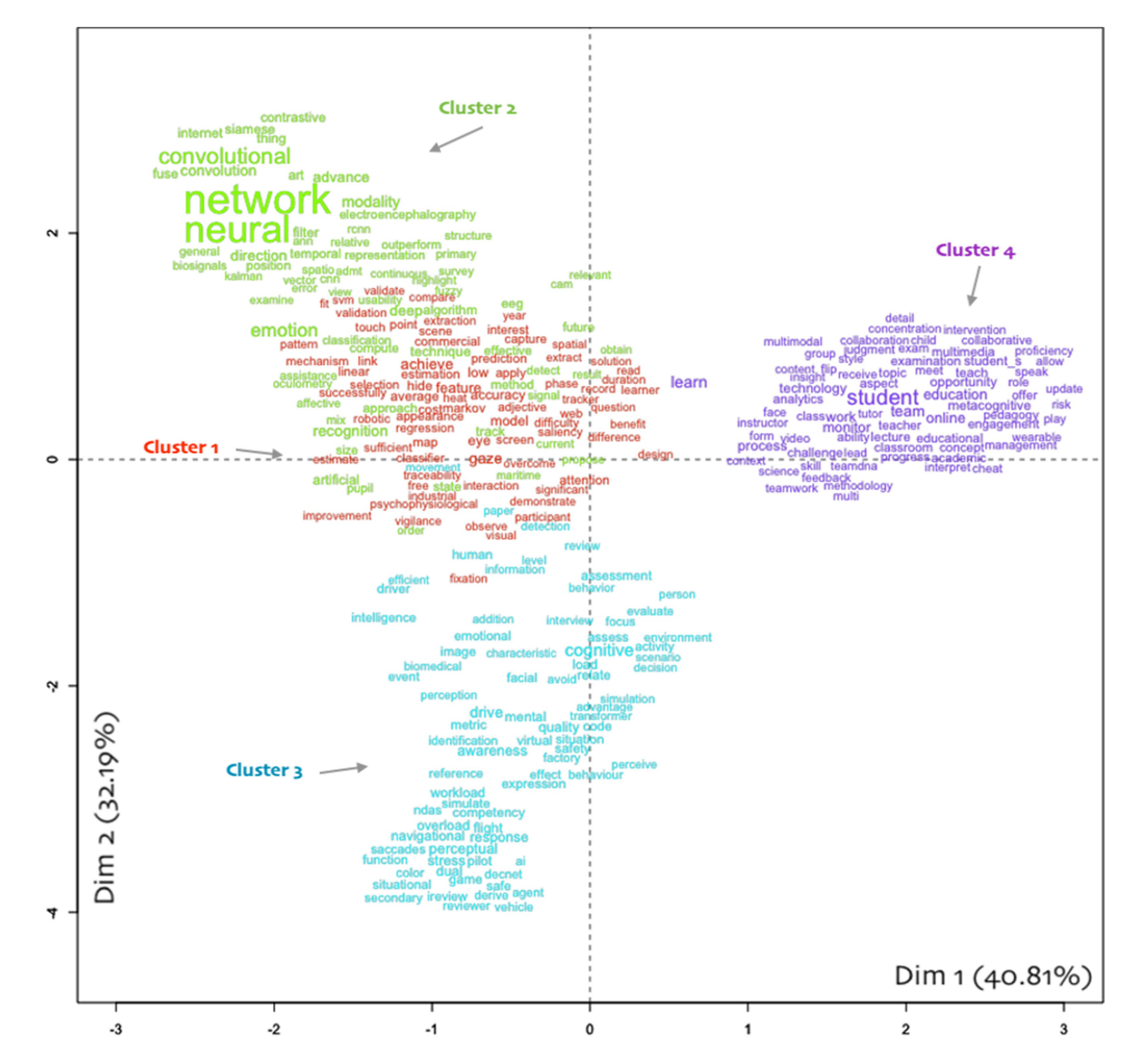


FIGURE 3
Correspondence analysis with factor 1 (40.81%) and factor 2 (32.19%).

context of robotic cars, Saha et al. (2018) proposed a CNN architecture that estimates the direction of vision from detected eyes and surpasses the latest results from the Eye-Chimera database. According to Rafee et al. (2022), previous eye-movement approaches focused on classifying eye movements into two categories: saccades and non-saccades. A limitation of these approaches is that they confuse fixations and smooth tracking by placing them in the non-saccadic category (Rafee et al., 2022). They proposed a low-cost optical motion analysis system with CCN technology and Kalman filters for estimating and analyzing the position of the eyes.

5.1.2. AIET to analyze and predict learners' behaviors

With this approach, engineering assessments in the age of AI take a new shift and offer diverse possibilities (Swiecki et al., 2022), especially with the increase of online education platforms and environments (Peng et al., 2022). As such, research suggests that machine learning technologies can provide better detection than current state-of-the-art event detection algorithms and achieve manual encoding performance (Zemblys et al., 2018). When applied in engineering education, AIET-based approaches have the potential to provide automatic and non-intrusive assessment (Meza et al., 2017; Ahrens, 2020; Chen, 2021), higher accuracy (Hijazi et al., 2021), complex dynamic scenes such as video-based data (Guo et al., 2022), and a less consuming process. For instance, Hijazi et al. (2021) used iReview, an intelligent tool for evaluating code review quality using biometric measures gathered from code reviewers (often called biofeedback).

Costescu et al. (2019) combined GP3 Eye Tracker with OGAMA to identify learners at risk of developing attention problems. They were able to accurately assess visual attention skills, interpret data, and predict reading abilities. Ahrens (2020) tracked how software

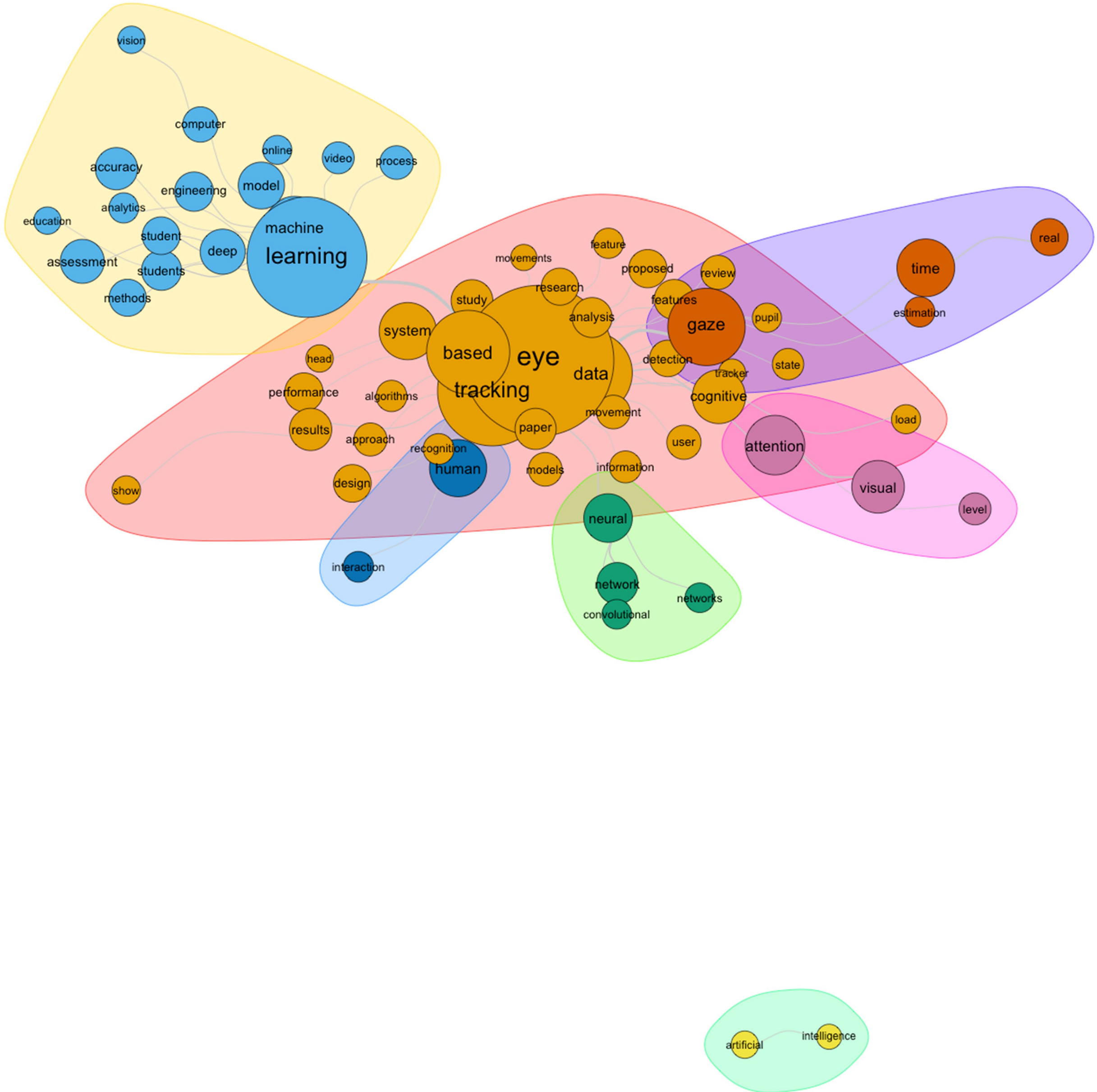


FIGURE 4
Graph of similarity of the whole corpus (title-abstract-keywords).

engineers navigate and interact with documents. By analyzing their areas of focus and gaze recordings, the author developed an algorithm to identify trace links between artifacts from these data. He finally concluded that eye tracking and interaction data are automatic and non-intrusive, allowing automatic recording without manual effort. This approach has interesting applications and perspectives for engineering design, namely the assessment of student visual parameters and algorithm replication for mass assessment, fairness and accuracy (objectivity, overload, and increased perception), understanding student learning behaviors, etc.

Moreover, as reported in our Review, AI and subsets for eye-tracking studies appear to be effective, as an average accuracy of approximately 80% was found for applications in engineering education, including in-class, VR, laboratory, and industrial settings.

5.2. Dimensions in engineering competency assessment of/for learning

With regard to our second research question, different dimensions of student learning have been analyzed. A somewhat unsurprising result was the prevalence of assessing student cognitive state, as eye tracking indeed relates to learners' visual cues. As such, multiple studies can be found within the pertinent literature over the last few decades focusing on the assessment of cognitive states (e.g., Hayes et al., 2011) and visual cognition and perception (e.g., Gegenfurtner et al., 2013; Rayner et al., 2014). However, taken together and considering the sample size ($N = 76$ papers) reported over the last decade, this Review revealed that few studies in the field of engineering education have focused on AI- and eye-tracking-based assessment of student learning.

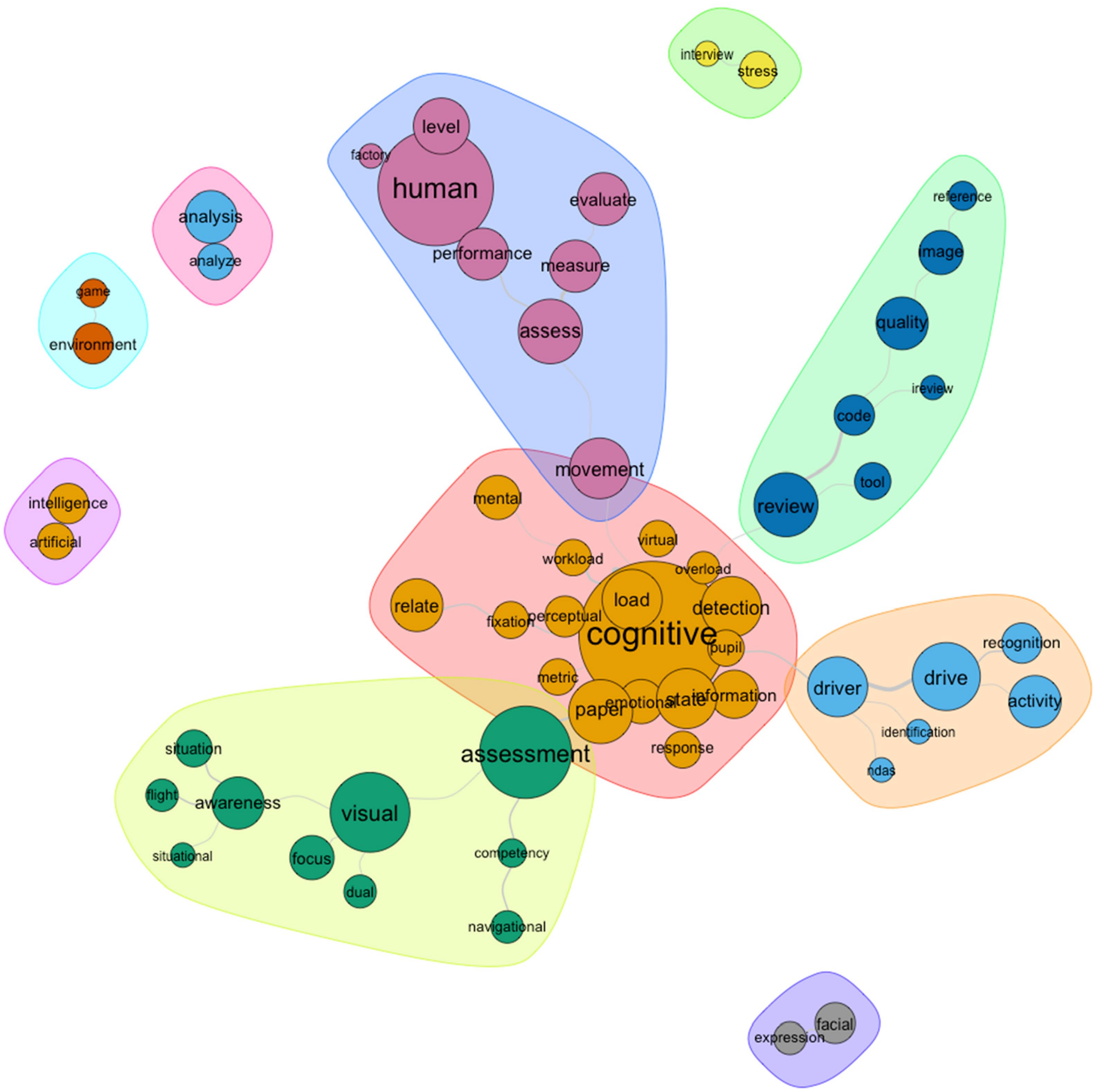


FIGURE 5
Cluster graph of "assessment" ($\chi^2 = 16.17$; $p < 0.0001$) and "assess" ($\chi^2 = 15.96$; $p < 0.0001$).

Although the expected finding was that papers would essentially focus on the visual and cognitive aspects of student learning and competencies, this study also shows an interest in the literature that focuses on other components, such as functional (skills) and social (attitudes) aspects of student learning. Indeed, it is assumed that eye tracking is essentially used as a tool for examining cognitive processes (Beesley et al., 2019). However, references for the meta competency aspect are seriously lacking. Several reasons may explain this repartition. First, it is true that early studies in this area focused primarily on obtaining insights into learners' visual patterns and therefore attempted to describe visual dynamics when learners look at the material in different environments and formats. Over the last decade, the focus has shifted to computational perspectives to visual attention modeling (e.g., Borji and Itti, 2012), driven by a digital transformation with the advances of attention computing, AI, machine learning, and cloud computing. Since 2013, and a bit later in 2016, as shown in Figure 7 (top), there has been a rapid rise of eye-tracking and AI-based assessments in research, especially when the field of AI becomes more accessible to cognitivists, psychologists, and engineering educational researchers. For instance, motivated by the complexity of contemporary visual materials and scenes, attention mechanism was associated with computer vision to imitate the human visual system (Guo et al., 2022). Moreover, this shift can be analyzed following the AI breakthroughs over the decade (2015: Russakovsky et al., 2015: OpenAI co-founded in 2015: deep learning models…). For instance, in January, 2023, the MIT Review published their 22nd 10 breakthrough technologies 2023 annual list (MIT Technology Review, 2023), recognizing key technological advances in many fields, such as AI. This list ranked "AI that makes images" in second position, justifying the growing interest visual computing has in contemporary research.

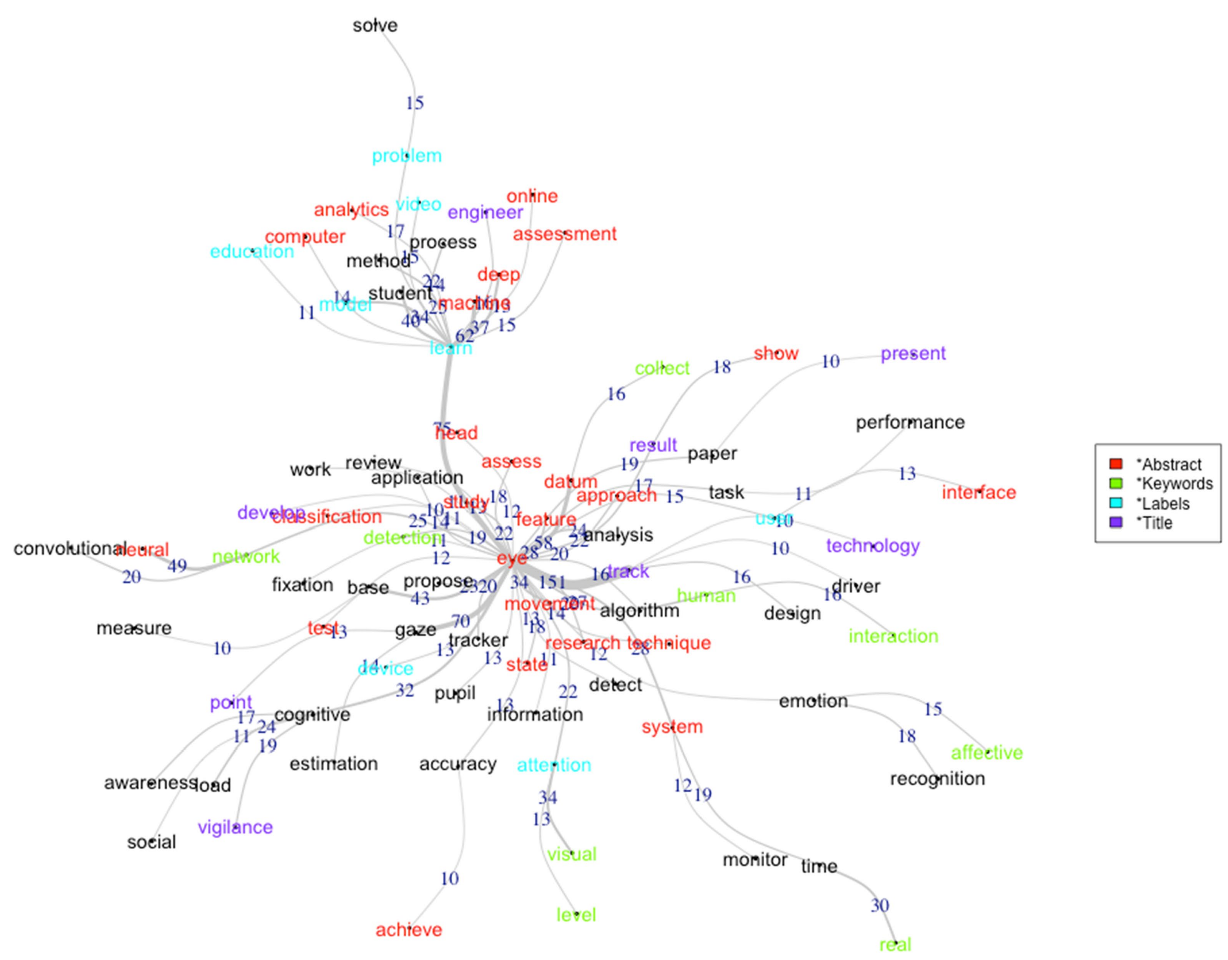


FIGURE 6
Graph of similarity of the corpus (tit-abs-key) with defined labels (addresses RQ2).

Shao et al. (2022) identified three waves of climax in AI advancements: in the early 60s, the second climax, and the third wave of AI, which according to LeCun et al. (2015) started with the era of deep learning, highly fostered developments and progress in society. As such, ImageNet was released in 2012, which in 2015 helped companies such as Microsoft and Google develop machines that could defeat humans in image recognition challenges. ImageNet was foundational to the advances of computer vison research (including recognition and visual computing).

We also reported the following different forms of assessments in engineering education: assessment of learning, i.e., as a summative evaluation (e.g., Bottos and Balasingam, 2020; Hijazi et al., 2021), formative assessment, i.e., assessment for learning, including feedback (e.g., Su et al., 2021; Tamim et al., 2021), self-assessment (Khosravi et al., 2022), and peer-assessment (Chen, 2021). In fact, engineering tasks are becoming increasingly complex. Therefore, current engineering instructions apply several assessments to better map student learning and their abilities, especially in active pedagogies such as project-based learning (PBL). This is reported by Ndiaye and Blessing (2023), who analyzed engineering instructors' course review reports and highlighted several combinations of assessment (e.g., summative: 2D project, exam, review, and prototype evaluation; formative: quizzes, problem sets, and homework assignment; peer assessment: peer review…). Providing an effective competency assessment for learning, especially feedback, to all students in such complex fields is challenging and time-consuming. Therefore, as there is a strong association between AI and ET, researchers have been exploring alternative solutions within this synergy. As such, Su et al. (2021) used video to analyze student concentration. They proposed a non-intrusive computer vision system based on deep learning to monitor students' concentration by extracting and inferring high-level visual signals of behavior, including facial expressions, gestures, and activities. A similar approach was used by Bottos and Balasingam (2020), who tracked reading progression using eye-gaze measurements and Hidden Markov models. With regard to team collaboration assessment, Guo and Barmaki (2020) used an automated tool based on gaze points and joint visual attention information from computer vision to assess team collaboration and cooperation.

5.3. Challenges of AIET

Despite the importance, AIET-based engineering assessment has some limitations. First, it suffers from a systematic and integrative approach of competency and competency assessment. Khan et al.

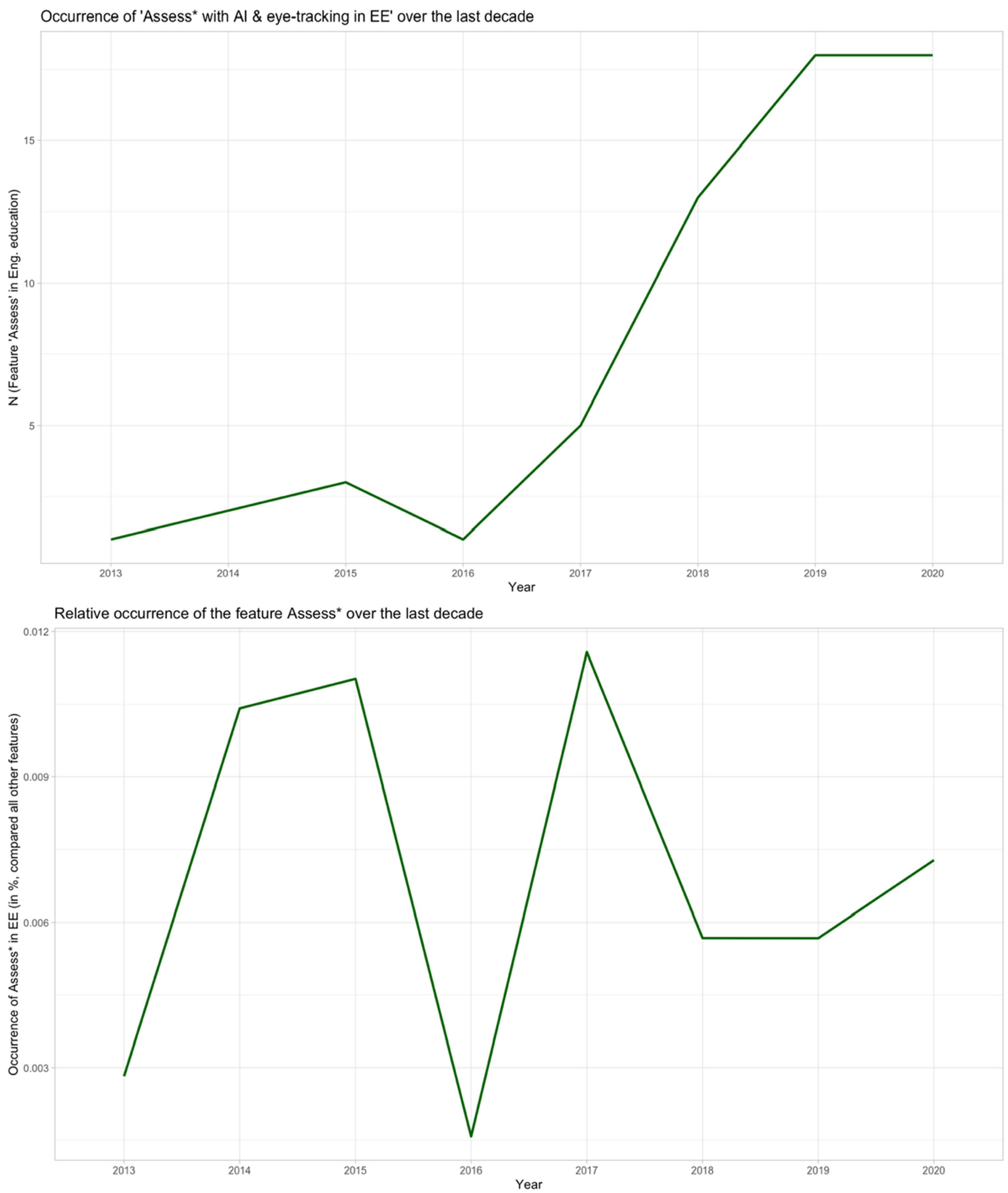


FIGURE 7
Absolute and relative occurrence of the feature "assess*" in the context of AIET.

(2023) reviewed the literature and identified a similar result for AI-based competency assessment in engineering design education. Indeed, competency, especially the measurement of student expertise, is viewed differently among researchers. There is ongoing debate about terminology within the literature (e.g., Le Deist and Winterton, 2005; Blömeke et al., 2015).

A second key challenge is the technique that is used to evaluate student learning. There are different eye-tracking methods and tools and they do not use the same tracking approach, hence not allowing tracking of the same behaviors. Consequently, further investigation is needed to achieve an appropriate network construction, followed by more efficient training to avoid common failures, such as over-training (e.g., Morozkin et al., 2017).

Other critical issues can be highlighted. AIET technologies are often too expensive and time-consuming (e.g., analysis of manual gaze data and data interpretation) to be implemented in

TABLE 5 Keyword-in-context (with 10 examples of a match).

[text7, 299] was developed to	| assess |	Vigilance levels
[text11, 128] neurophysiological approach to		Workers' stress
[text14, 216] data are used to		The workers' ergonomic performance
[text39, 15] tracking data to		Cognitive vigilance levels
[text40, 134] to measure and		Navigational competence
[text41, 60] data allows to		The cognitive load
[text43, 50] in order to		Virtual agent's eye
[text51, 73] we proposed to		The visualization environment
[text52, 185] are used to		The reviewer's comprehension
[text53, 85] able to accurately		Their visual attention

TABLE 6 Application scenarios.

Scenarios	Task focus	Application/Testing	References
Classroom learning	Student attention and engagement: use of ordinary web cam-based and computer vision algorithms to estimate (individually and in groups) and display student attention levels through easy color-coded charts for the instructor to take the necessary action during the lecture.	Classroom-based	Renawi et al. (2022)
	Self-directed learning environment: development of a low-cost webcam-based eye tracking solution combined with machine learning algorithms. The model implemented to a 4-min engineering lecture can achieve similar accuracy compared with the head-worn tracker.	Classroom-based: third year engineering students	Khosravi et al. (2022)
	Class insight, a student monitoring system: development of a machine learning-based monitoring system that allows teachers to submit an assessment to students in a completely paperless way. The system tracks students' faces and eyes during reading and updates the progress immediately, hence helping instructors to monitor tasks in real time.	Classroom-based	Tamim et al. (2021)
Simulation training: situational awareness	Flight simulation: a situation awareness (SA) assessment method based on an AI neural network (NN) and integrating visual cues and flight control is developed and resulted in 96% accuracy of the SA classification of the NN model to the experimental data set.	Simulated flight training experiments for flight cadets	Jiang et al. (2022)
	Navigational competency: development of an AI-based competency assessment tool for safe navigation (AICATSAN) for various behaviors, such as situational awareness, decision making, teamwork, and communication and influencing skills.	Maritime navigational safety	Lili et al. (2021)
Lab practice	Human-machine interaction: characterization performed on two types of eye tracking devices to support the development of cognitive human-machine systems.	Laboratory	Lim et al. (2019)
VR/AR settings	Cognitive load assessment: proposition of an autonomous, privacy- preserving, and attention-based cognitive load recognition system for drivers under critical conditions based on driving data collected from a previously simulated VR driving environment. Multiple classifiers were trained to help assess the driver's cognitive load. Integrating the visual ET data into the VR configurations improves the accuracy (>80%) to predict user cognitive load.	User interface	Bozkir et al. (2019)
Other industry settings	Software traceability: development of an algorithm aiming to track how software engineers interact with documents and record eye connections between these documents.	Document interaction in industry	Ahrens (2020)

classroom practice. Therefore, the development of low-cost approaches can be a better and more inclusive approach for engineering learning and instructions. Finally, the assessment of the student (behavior, mental state, etc.) often tends to replace the assessment of student learning (outcomes). It is not clear how studies clarify this difference.

6. Conclusion

This Review provides important insights into AI- and eye-tracking-based competency assessment in engineering education. With regard to our first research question (RQ1), this Review revealed that research trends have taken two orientations over the last decade.

TABLE 7 Accuracy of AI and ET in engineering assessment.

References	Topic	Accuracy (%)
Jiang et al. (2022)	Awareness in flight simulation	96
Wu et al. (2021)	Emotion classification on ET and EEG fused signals employing deep gradient neural networks	88.1
Li et al. (2021)	Predicting the level of difficulty in spatial visualization problems	87.6
Xin et al. (2021)	Detecting the difficulty of the task	91.8
Shi et al. (2020)	A neurophysiological approach to assess training outcomes under stress: a VR experiment	80.38
Bottos and Balasingam (2020)	Tracking the progression of reading using eye-gaze point measurements and hidden Markov models	27.1
Bharadva (2021)	An ML approach to detect student online engagement	88.9
Singh and Modi (2022)	A camera-based eye gaze tracking system to analyze visual attention using deep learning	84
Chen (2021)	Automatic measurement of teamwork processes	75
Farha et al. (2021)	Assessment of cognitive vigilance levels	76.8
Bozkir et al. (2019)	Autonomous and real-time assessment of cognitive load using ET in a VR setup	80
Pritalia et al. (2020)	Classification of learning approaches in multimedia learning using ET and ML	71
Hijazi et al. (2021)	A code evaluation tool using biofeedback (iReview)	87
Chakraborty et al. (2021)	A human-robot system estimating the visual focus of the attention level	99.43
Singh et al. (2018)	Guiding the selection of software inspectors	94
Bautista and Naval (2021)	CLRGaze: representations of eye movement signals	97.3
Gite et al. (2021)	ADMT: driver motion tracking system	12
Aunsri and Rattarom (2022)	Eye-based features for head-free gaze estimation using web cameras	97.71

We showed that research generally discussed that (1) eye-tracking devices developed intrinsically with AI to enhance the gaze-tracking process (improvement of techniques), and/or (2) AI can be used to analyze, predict, and assess eye-tracking analytics (application domain). With regard to RQ2, i.e., the salient competency dimensions and labels attributed to assessing engineering education, the main finding is that visual cognitive aspects of learner competency are a primary focus. Hence, despite growing interest in advanced technologies, such as AI, attention computing, and eye-tracking, it is shown that student competency and underlying components are assessed in a fragmented way, i.e., not in a systematic and integrative approach to engineering competency and holistic assessment. Assessing engineering student expertise with AIET is essentially limited to visual aspects, and there is a lack of references and understanding about how it can be extended to more complex engineering learning. Therefore, we argue that such limitations can be situated in the technology itself, which relies on the eye (hence visual cognition and perception only) as a portal to an individual brain to understand human behavior. In addition, there is not yet a common understanding of expertise and competency. Terminologies vary depending on the subject domain.

This Review presents some limitations. Although the debate about competency or competence is still ongoing within the literature, we focus on engineering competency in terms of dimensions to analyze what is being effectively assessed. However, as preliminary research, an approach may need to be extended to other underlying engineering fields and explore different possible components in student competency acquisition. This needs to be better clarified with regard to existing frameworks. Additionally, as for every review, we only used well known terms; however, many terminologies are being used to describe eye-tracking techniques and studies (eye or gaze tracking, eye movements, visual tracking, etc.), including the variation in the syntax of the words (e.g., eye tracking or eye tracking or eye tracking) and competency (competence, ability, etc.). AI also suffers from this variation (e.g., machine learning, deep learning, NLP, etc.). Not all these terms were used, thus reducing the search.

This Review is probably one of the first to discuss trends in research on the assessment of engineering education with AIET technologies. Multiple relevant perspectives are possible. For engineering education, it is important to investigate in-depth how AIET can support complex learning and instruction. AIET may open new opportunities to better assess learning inclusively and efficiently, assuming that relevant assessment frameworks of the content to be assessed are well defined and situated. It is necessary to examine the combination of holistic approaches to assess complex engineering skills. As such, this Review may have several implications for the integration of AIET in engineering education. It may open new research perspectives on the.

AIET-based assessment of student learning, which will be worth investigating. This is a key area to be explored in-depth further.

Future research can focus on exploring multimodal approaches to better capture less-represented dimensions of engineering student competencies, helping to mitigate existing assessment shortcomings. One of the main issues is mapping student abilities and their engagement holistically during their learning with different assessments methods. Therefore, an increasing interest lies in associating different inclusive fine-grained techniques, such as electrical (EEG), physiological (heart-rate variability, galvanic skin resistance, and eye tracking), neurophysiological (fMRI) signals, and other traditional assessments (e.g., self-reported surveys, quizzes, peer-assessment, etc.), to improve assessment accuracy and efficiency. For instance, Wu et al. (2021) developed a deep-gradient neural network for the classification of multimodal signals (EEG and ET). Their model predicted emotions with 81.10% accuracy during an experiment with eight emotion event stimuli. Similar studies exploring learning and assessment are needed to gain holistic insights into student learning, instructions, and assessments.

Author contributions

YN and KL made a substantial, direct, and intellectual contribution to the work. LB reviewed and provided general comments on the work. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Ministry of Education, Singapore (MOE2020-TRF-029).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahrens, M. (2020). "Towards automatic capturing of traceability links by combining eye tracking and interaction data" In *2020 IEEE 28th International Requirements Engineering Conference (RE)*, Zurich, Switzerland. 434–439.

Amadori, P. V., Fischer, T., Wang, R., and Demiris, Y. (2021). Predicting secondary task performance: a directly actionable metric for cognitive overload detection. *IEEE Transac. Cogn. Dev. Syst.* 14, 1474–1485. doi: 10.1109/TCDS.2021.3114162

Amri, S., Ltifi, H., and Ayed, M. B. (2017). "Comprehensive evaluation method of visual analytics tools based on fuzzy theory and artificial neural network" In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 1–6. IEEE.

Aracena, C., Basterrech, S., Snáel, V., and Velásquez, J. (2015). "Neural networks for emotion recognition based on eye tracking data" In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2632–2637, IEEE.

Aunsri, N., and Rattarom, S. (2022). Novel eye-based features for head pose-free gaze estimation with web camera: new model and low-cost device. *Ain Shams Eng. J.* 13:101731. doi: 10.1016/j.asej.2022.101731

Bautista, L. G. C., and Naval, P. C. (2021) "CLRGaze: contrastive learning of representations for eye movement signals, 2021" In *29th European Signal Processing Conference (EUSIPCO)*, Dublin, Ireland. 1241–1245.

Beesley, T., Pearson, D., and Le Pelley, M. (2019). "Eye tracking as a tool for examining cognitive processes" in *Biophysical measurement in experimental social science research*. ed. G. Foster (Cambridge, MA: Academic Press), 1–30.

Bharadva, K. R. (2021). A machine learning framework to detect student's online engagement. Doctoral dissertation. Dublin, National College of Ireland.

Blömeke, S., Gustafsson, J.-E., and Shavelson, R. J. (2015). Beyond dichotomies: competence viewed as a continuum. *Z. Psychol.* 223, 3–13. doi: 10.1027/2151-2604/a000194

Borji, A., and Itti, L. (2012). State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 185–207. doi: 10.1109/TPAMI.2012.89

Bottos, S., and Balasingam, B. (2020). Tracking the progression of reading using eye-gaze point measurements and hidden markov models. *IEEE Trans. Instrum. Meas.* 69, 7857–7868. doi: 10.1109/TIM.2020.2983525

Bozkir, E., Geisler, D., and Kasneci, E. (2019). "Person independent, privacy preserving, and real time assessment of cognitive load using eye tracking in a virtual reality setup" In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan. 2019, 1834–1837.

Chakraborty, P., Ahmed, S., Yousuf, M. A., Azad, A., Alyami, S. A., and Moni, M. A. (2021). A human-robot interaction system calculating visual focus of human's attention level. *IEEE Access* 9, 93409–93421. doi: 10.1109/ACCESS.2021.3091642

Chen, X. (2021). TeamDNA: Automatic measures of effective teamwork processes from unconstrained team meeting recordings. Doctoral dissertation. Rice University.

Costescu, C., Rosan, A., Brigitta, N., Hathazi, A., Kovari, A., Katona, J., et al. (2019). "Assessing visual attention in children using GP3 eye tracker" In *10th IEEE International Conference on Cognitive Info Communications (CogInfoCom)*, Naples, Italy. 343–348.

Das, A., and Hasan, M. M. (2014). "Eye gaze behavior of virtual agent in gaming environment by using artificial intelligence" In *IEEE International Conference on Electrical Information and Communication Technology (EICT)*, 1–7.

Dogan, K. M., Suzuki, H., and Gunpinar, E. (2018). Eye tracking for screening design parameters in adjective-based design of yacht hull. *Ocean Eng.* 166, 262–277. doi: 10.1016/j.oceaneng.2018.08.026

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). "An image is worth 16x16 words: transformers for image recognition at scale" In *International Conference on Learning Representations*.

Duchowski, A., and Duchowski, A. (2007). "Eye tracking techniques" in *Eye tracking methodology: Theory and practice*. ed. A. T. Duchowski (London: Springer), 51–59.

Farha, N. A., Al-Shargie, F., Tariq, U., and Al-Nashash, H. (2021). "Artifact removal of eye tracking data for the assessment of cognitive vigilance levels" In *IEEE Sixth International Conference on Advances in Biomedical Engineering (ICABME)*, 175–179.

Frerejean, J., van Merriënboer, J. J., Kirschner, P. A., Roex, A., Aertgeerts, B., and Marcellis, M. (2019). Designing instruction for complex learning: 4C/ID in higher education. *Eur. J. Dent. Educ.* 54, 513–524. doi: 10.1111/ejed.12363

Gegenfurtner, A., Lehtinen, E., and Säljö, R. (2013). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educ. Psychol. Rev.* 23, 523–552. doi: 10.1007/s10648-011-9174-7

Gite, S., Pradhan, B., Alamri, A., and Kotecha, K. (2021). ADMT: advanced driver's movement tracking system using spatio-temporal interest points and maneuver anticipation using deep neural networks. *IEEE Access* 9, 99312–99326. doi: 10.1109/ACCESS.2021.3096032

Guo, Z., and Barmaki, R. (2020). Deep neural networks for collaborative learning analytics: evaluating team collaborations using student gaze point prediction. *Australas. J. Educ. Technol.* 36, 53–71. doi: 10.14742/ajet.6436

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., et al. (2022). Attention mechanisms in computer vision: a survey. *Computat. Visual Media* 8, 331–368. doi: 10.1007/s41095-022-0271-y

Hadgraft, R. G., and Kolmos, A. (2020). Emerging learning environments in engineering education. *Australas. J. Eng. Educ.* 25, 3–16. doi: 10.1080/22054952.2020.1713522

Han, Y., Han, B., and Gao, X. (2021). Human scanpath estimation based on semantic segmentation guided by common eye fixation behaviors. *Neurocomputing* 453, 705–717. doi: 10.1016/j.neucom.2020.07.121

Hayes, T. R., Petrov, A. A., and Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on raven's advanced progressive matrices. *J. Vis.* 11, 1–11. doi: 10.1167/11.10.10

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition" In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

Hijazi, H., Cruz, J., Castelhano, J., Couceiro, R., Castelo-Branco, M., de Carvalho, P., et al. (2021). "iReview: an intelligent code review evaluation tool using biofeedback" In *2021 IEEE 32nd International Symposium on Software Reliability Engineering (ISSRE)*, Wuhan, China. 476–485.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks" In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Jiang, G., Chen, H., Wang, C., and Xue, P. (2022). Transformer network intelligent flight situation awareness assessment based on pilot visual gaze and operation behavior data. *Int. J. Pattern Recognit. Artif. Intell.* 36:2259015. doi: 10.1142/S0218001422590157

Khan, S., Blessing, L., and Ndiaye, Y. (2023). "Artificial intelligence for competency assessment in design education: a review of literature" In *9th International Conference on Research Into Design*, Indian Institute of Science, Bangalore, India.

Khosravi, S., Khan, A. R., Zoha, A., and Ghannam, R. (2022) "Self-directed learning using eye-tracking: a comparison between wearable head-worn and webcam-based technologies" In *IEEE Global Engineering Education Conference (EDUCON)*, Tunis, Tunisia, 640–643.

Kiefer, P., Giannopoulos, I., Raubal, M., and Duchowski, A. (2017). Eye tracking for spatial research: cognition, computation, challenges. *Spat. Cogn. Comput.* 17, 1–19. doi: 10.1080/13875868.2016.1254634

Le Deist, F. D., and Winterton, J. (2005). What is competence? *Hum. Resour. Dev. Int.* 8, 27–46. doi: 10.1080/1367886042000338227

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, Z., Guo, P., and Song, C. (2021). A review of main eye movement tracking methods. *J. Phys.* 1802:042066. doi: 10.1088/1742-6596/1802/4/042066

Li, X., Younes, R., Bairaktarova, D., and Guo, Q. (2020). Predicting spatial visualization problems' difficulty level from eye-tracking data. *Sensors* 20:1949. doi: 10.3390/s20071949

Lili, L., Singh, V. S., Hon, S., Le, D., Tan, K., Zhang, D., et al. (2021). "AI-based behavioral competency assessment tool to enhance navigational safety" In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. Mauritius. 2021, 1–6.

Lim, Y., Gardi, A., Pongsakornsathien, N., Sabatini, R., Ezer, N., and Kistan, T. (2019). Experimental characterisation of eye-tracking sensors for adaptive human-machine systems. *Measurement* 140, 151–160. doi: 10.1016/j.measurement.2019.03.032

Mehta, P., Malviya, M., McComb, C., Manogharan, G., and Berdanier, C. G. P. (2020). Mining design heuristics for additive manufacturing via eye-tracking methods and hidden markov modeling. *ASME. J. Mech. Des.* 142:124502. doi: 10.1115/1.4048410

Meza, M., Kosir, J., Strle, G., and Kosir, A. (2017). Towards automatic real-time estimation of observed learner's attention using psychophysiological and affective signals: the touch-typing study case. *IEEE Access* 5, 27043–27060. doi: 10.1109/ACCESS.2017.2750758

MIT Technology Review (2023). 10 breakthrough technologies 2023, MIT technology review. Available at: https://www.technologyreview.com/2023/01/09/1066394/10-breakthrough-technologies-2023/#crispr-for-high-cholesterol (Accessed 19 August 2023).

Morozkin, P., Swynghedauw, M., and Trocan, M. (2017). "Neural network based eye tracking" in *Computational collective intelligence. ICCCI 2017 Lecture Notes in Computer Science*. eds. N. Nguyen, G. Papadopoulos, P. Jędrzejowicz, B. Trawiński and G. Vossen (Cham: Springer), 10449.

Muldner, K., and Burleson, W. (2015). Utilizing sensor data to model students' creativity in a digital environment. *Comput. Hum. Behav.* 42, 127–137. doi: 10.1016/j.chb.2013.10.060

Ndiaye, Y., and Blessing, L. (2023). Assessing performance in engineering design education from a multidisciplinary perspective: an analysis of instructors' course review reports. *Proc. Design Soc.* 3, 667–676. doi: 10.1017/pds.2023.67

Peng, C., Zhou, X., and Liu, S. (2022). An introduction to artificial intelligence and machine learning for online education. *Mobile Netw. Appl.* 27, 1147–1150. doi: 10.1007/s11036-022-01953-3

Płużyczka, M. (2018). The first hundred years: a history of eye tracking as a research method. *Appl. Linguist. Papers* 4, 101–116. doi: 10.32612/uw.25449354.2018.4.pp.101-116

Pritalia, G. L., Wibirama, S., Adji, T. B., and Kusrohmaniah, S. (2020). "Classification of learning styles in multimedia learning using eye-tracking and machine learning" In *2020 FORTEI-International Conference on Electrical Engineering (FORTEI-ICEE)*, 145–150, IEEE.

Rafee, S., Xu, Y., Zhang, J. X., and Yemeni, Z. (2022). Eye-movement analysis and prediction using deep learning techniques and Kalman filter. *Int. J. Adv. Comput. Sci. Appl.* 13, 937–949. doi: 10.14569/IJACSA.2022.01304107

Ratinaud, P. (2009). IRAMUTEQ: Interface de R pour les analyses multidimensionnelles de textes et de questionnaires. Available at: http://www.iramuteq.org

Rayner, K., Loschky, L. C., and Reingold, E. M. (2014). Eye movements in visual cognition: the contributions of George W. *McConkie. Vis. Cogn.* 22, 239–241. doi: 10.1080/13506285.2014.895463

Reinert, M. (1990). ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. *Bull. Méthodol. Sociol.* 26, 24–54. doi: 10.1177/075910639002600103

Renawi, A., Alnajjar, F., Parambil, M., Trabelsi, Z., Gochoo, M., Khalid, S., et al. (2022). A simplified real-time camera-based attention assessment system for classrooms: pilot study. *Educ. Inf. Technol.* 27, 4753–4770. doi: 10.1007/s10639-021-10808-5

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Saha, D., Ferdoushi, M., Emrose, M. T., Das, S., Hasan, S. M., Khan, A. I., et al. (2018) "Deep learning-based eye gaze controlled robotic car" In *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE. pp. 1–6.

Shao, Z., Zhao, R., Yuan, S., Ding, M., and Wang, Y. (2022). Tracing the evolution of AI in the past decade and forecasting the emerging trends. *Expert Syst. Appl.* 209:118221. doi: 10.1016/j.eswa.2022.118221

Shavelson, R. J. (2013). On an approach to testing and modeling competence. *Educ. Psychol.* 48, 73–86. doi: 10.1080/00461520.2013.779483

Shi, Y., Zhu, Y., Mehta, R. K., and Du, J. (2020). A neurophysiological approach to assess training outcome under stress: a virtual reality experiment of industrial shutdown maintenance using functional near-infrared spectroscopy (fNIRS). *Adv. Eng. Inform.* 46:101153. doi: 10.1016/j.aei.2020.101153

Siddique, Z., Panchal, J., Schaefer, D., Haroon, S., Allen, J. K., and Mistree, F. (2012). "Competencies for innovating in the 21st century," In *International design engineering technical conferences and computers and information in engineering conference*. American Society of Mechanical Engineers. 45066, 185–196.

Singh, A., and Das, S. (2022). "A cheating detection system in online examinations based on the analysis of eye-gaze and head-pose" In *THEETAS 2022: Proceedings of The International Conference on Emerging Trends in Artificial Intelligence and Smart Systems, THEETAS 2022*. 16–17 April 2022, Jabalpur, India, p. 55, European Alliance for Innovation.

Singh, J., and Modi, N. (2022). A robust, real-time camera-based eye gaze tracking system to analyze users' visual attention using deep learning. *Interact. Learn. Environ.* 1–22:561. doi: 10.1080/10494820.2022.2088561

Singh, M., Walia, G. S., and Goswami, A. (2018) Using supervised learning to guide the selection of software inspectors in industry. In: *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Memphis, TN, USA*, pp. 12–17.

Spencer, L. M. (1997). "Competency assessment methods" in *Assessment, development, and measurement*. eds. L. J. Bassi and D. F. Russ-Eft, vol. *1* (Alexandria, VA: American Society for Training & Development)

Su, M. C., Cheng, C. T., Chang, M. C., and Hsieh, Y. Z. (2021). A video analytic in-class student concentration monitoring system. *IEEE Trans. Consum. Electron.* 67, 294–304. doi: 10.1109/TCE.2021.3126877

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., et al. (2022). Assessment in the age of artificial intelligence. *Comput. Educ. Artif. Intellig.* 3:100075. doi: 10.1016/j.caeai.2022.100075

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions" In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

Tamim, H., Sultana, F., Tasneem, N., Marzan, Y., and Khan, M. M. (2021) "Class insight: a student monitoring system with real-time updates using face detection and eye tracking" In *IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, 2021, pp. 0213–0220.

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G. Z., and Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. *J. Surg. Res.* 191, 169–178. doi: 10.1016/j.jss.2014.04.032

van Merriënboer, J. J. G., and Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design*. *3rd* Edn. New York and London: Routledge.

Wade, N., and Tatler, B. W. (2005) *The moving tablet of the eye: The origins of modern eye movement research*. USA: Oxford University Press

Wang, X., Zhao, X., and Zhang, Y. (2021). Deep-learning-based reading eye-movement analysis for aiding biometric recognition. *Neurocomputing* 444, 390–398. doi: 10.1016/j.neucom.2020.06.137

Wu, Q., Dey, N., Shi, F., Crespo, R. G., and Sherratt, R. S. (2021). Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks. *Appl. Soft Comput.* 110:107752. doi: 10.1016/j.asoc.2021.107752

Xin, L., Bin, Z., Xiaoqin, D., Wenjing, H., Yuandong, L., Jinyu, Z., et al. (2021). Detecting task difficulty of learners in colonoscopy: evidence from eye-tracking. *J. Eye Mov. Res.* 14:10. doi: 10.16910/jemr.14.2.5

Zemblys, R., Niehorster, D. C., Komogortsev, O., and Holmqvist, K. (2018). Using machine learning to detect events in eye-tracking data. *Behav. Res. Ther.* 50, 160–181. doi: 10.3758/s13428-017-0860-3

Zlatkin-Troitschanskaia, O., and Pant, H. A. (2016). Measurement advances and challenges in competency assessment in higher education. *J. Educ. Meas.* 53, 253–264. doi: 10.1111/jedm.12118

Frontiers in
Education

Explores education and its importance for individuals and society

A multidisciplinary journal that explores research-based approaches to education for human development. It focuses on the global challenges and opportunities education faces, ultimately aiming to improve educational outcomes.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact