

Eye tracking in physics education research: A systematic literature reviewL. Hahn^{*} and P. Klein[†]*Faculty of Physics Education Research, University of Göttingen,
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany* (Received 6 October 2021; accepted 22 December 2021; published 2 March 2022)

Eye tracking is becoming increasingly popular in physics education research (PER). As technology has advanced considerably in recent years and has become more user friendly, it is anticipated that eye tracking will play an increasingly significant role in assessing student learning at the process level in future studies. The main objective of this systematic review is to summarize the current status quo regarding eye tracking in PER and reviewing (a) the dissemination, (b) the methodological implementation, and (c) the insights provided by eye tracking in PER. We identified 33 journal articles, published between March 2005 and April 2021, that used eye tracking for original empirical research in the area of physics education. The results reveal that although eye tracking has been used in many different areas of physics, a clear focus on mechanics is evident, particularly for measuring visual attention in assessment scenarios like problem solving. While a high methodological rigor in the selection and analysis of the visual stimuli was apparent, only a few studies have provided a complete documentation of the technological implementation (e.g., movement restrictions, accuracy, and calibration information) and a theoretical embedding for interpreting eye-tracking data. To synthesize the results of the different studies, we created an inductive category system in accordance with the considered independent variables of the studies. Accordingly, visual attention was most frequently compared between levels of performance (correct vs incorrect or high vs low achievers), thereby leading to performance-discriminating factors of eye movement across studies. Furthermore, learners' eye movements were compared across different stimuli, different time points, or between student groups to inform multimedia design and shed light on students' learning progression. In summary, eye tracking is particularly useful for studying processes in different domains that are relevant to PER. Specific gaps in the literature, methodological limitations, and implications of existing findings were also identified to recommend future research and practices.

DOI: [10.1103/PhysRevPhysEducRes.18.013102](https://doi.org/10.1103/PhysRevPhysEducRes.18.013102)**I. INTRODUCTION**

The eyes play a crucial role in information intake. Eye movements reveal information regarding a person's visual attention—that is, the location where a person's eyes are fixated. Eye tracking enables researchers to study the allocation of visual attention, thereby implying it can tell us which elements of a visual stimulus (e.g., learning material) a participant looked at, for how long, and in what order. Observing eye movements and shifts in visual attention among different regions can potentially provide valuable insights into cognitive processing. For this reason, eye-tracking studies have been conducted by physics education researchers for over a decade to gain insights

into students' learning processes or problem-solving strategies. What is unique about using eye tracking in physics education research (PER) is the challenge of investigating disciplinary concepts as opposed to other fields, such as multimedia research and cognitive science. As Rouinfar *et al.* put it, “physics problem solving is among the most intellectually and cognitively demanding processes that human beings are capable of engaging in” [1] (p. 1). Hence, it is not clear *per se* which educational topics in physics can be addressed by eye tracking or whether learning processes in physics can be informed by something as mundane as the perceptual functions involved in visual attention at all. Against this background, it is useful to have an overview of the research conducted in this field thus far.

Indeed, contributions to understanding students' eye movements are noticeably increasing, particularly in the journal *Physical Review Physics Education Research* [2]. Owing to commercial systems that are easy to use, and due to ample basic literature on methodological implementation and theoretical connections (e.g., Refs. [3,4]), researchers can enter this field of research with increasing

^{*}larissa.hahn@uni-goettingen.de[†]pascal.klein@uni-goettingen.de

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

ease. Having an overview helps connecting one's own work to the emerging field of eye tracking, while showing how various methods have been used in the PER community thus far.

While four reviews summarized eye-tracking research in the areas of expertise research [5], learning processes [6], multimedia learning [7], and mathematics education [8], to the best of our knowledge, there is still no systematic review of eye tracking in PER; moreover, existing reviews are not characteristic of PER. Therefore, in light of the above arguments, this article aims to fill this gap by disclosing the current inventory of where and how eye tracking has been used in PER thus far and what added value the method might offer to the field of PER.

II. GOALS AND RESEARCH QUESTIONS

Conducting a systematic literature review that includes all studies that use eye tracking in PER can have several purposes. First, an overview would provide insight into PER-specific approaches to use eye tracking as a research method that can be related to and opposed to general approaches in other education research practices. Second, in addition to technical, bibliographical, and methodological aspects, the educational impact of the use of eye tracking will be considered. More specifically, we intend to reveal the extent to which the use of eye tracking can provide added value for PER. Third, the literature review indicates the potential for further studies and points towards research topics that have not been exploited yet, hence providing (beginning) researchers with guidance in this regard. Fourth, as this review concentrates on eye tracking, the common practices and established procedures become evident by extracting documentation on the technological implementation. Thus, the following research questions are addressed:

RQ1: How widespread is the eye-tracking method in PER?

The first research question addresses the dissemination of the method. This includes identifying the number of PER studies that have used eye tracking thus far, when and where they were published, and which learning scenarios were investigated therein. We also address the (domain-specific) subject areas that have been investigated.

RQ2: What experimental setups have been used in PER-related eye-tracking studies?

The second research question characterizes the methodological implementation of eye tracking in PER—including analysis methods of visual stimuli, participants, technical equipment and procedure—and additional data.

RQ3: How was the eye-movement data interpreted, what results were obtained, and what are the implications of eye-tracking studies for PER?

Finally, we also identify the insights that eye tracking has provided to the field of PER. Therefore, we extract the reported eye-tracking metrics and the other variables that

were collected in extant studies. We summarize the relevant findings on visual attention that are important for PER. In addition, we analyze how the results have been interpreted and theoretically framed, and present the limitations that have been mentioned in the studies.

III. METHOD

With this systematic review, we aim to comprehensively summarize the existing literature pertaining to our research questions. To do this, we examine the studies with the highest quality and those that are most up to date and that are characterized by a rigorous peer-review process for quality assurance. Therefore, we decided to include only journal articles to answer the research questions. However, being comprehensive requires finding as many published studies as possible, including those that may only be progress reports described in proceedings or book chapters. In addition, to provide a comprehensive picture of the use of eye tracking in PER, we identified other sources through a snowball search, but only included them with respect to the first research question (dissemination) and the task type of the stimuli (RQ2). Mainly due to their brevity, studies published as proceedings or book chapters do not contain sufficient information for our research questions to assess the design, methods, results, and outcomes of the reported studies. This applies, for example, to one of the most important conference proceedings in PER—that are, the PERC proceedings, which are limited to four pages.

To sum it up, our search strategy reflects a trade-off between quality and completeness of coverage. To ensure a high level of transparency in the selection process and to improve the reporting of our review, the PRISMA instructions have been incorporated into the methodology [9].

A. Search strategy

As the first step (see Fig. 1, “Identification”), the *Web of Science* database was selected to access the relevant research studies in this review by logically linking variants of the following keywords using the AND Boolean operator: “eye tracking,” “science or physics,” and “learning-related terms.” The exact string reads (“eye gaze” OR “eye movement” OR “eye movements” OR “eye-movement” OR “eye-movements” OR “eyetracking” OR “eye tracking” OR “eye-tracking” OR “eye gaze tracking” OR “eye-gaze tracking” OR “eye-based gaze tracking” OR “eyegaze tracking” OR “visual attention”) AND (“science” OR “physics”) AND (“education” OR “instruction” OR “teaching” OR “learning” OR “problem solving” OR “students”). Searching this database ensures that the highest quality and most up-to-date research can be examined by limiting the literature search to those studies that have been published as articles in peer-reviewed journals (Fig. 1, top level, left).

After applying the inclusion criteria (see next section), the bibliographies of the remaining articles were screened

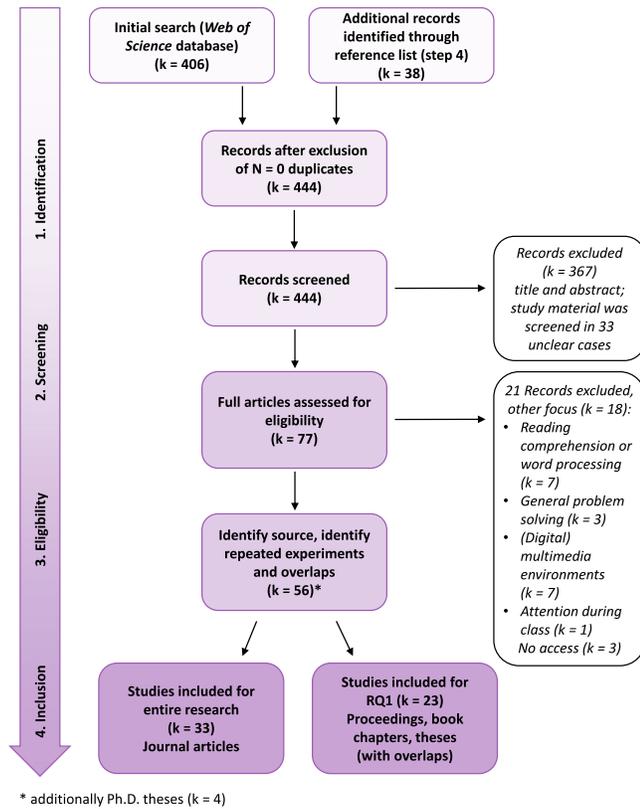


FIG. 1. Selection process of finding studies related to the research questions.

and a reverse search was conducted using GoogleScholar (i.e., all articles citing the remaining articles were identified). Qualifying journal articles were included in the search at the same level (Fig. 1, top level, right). In this step, studies from other sources (proceedings, book chapters, theses) that refer to the journal articles were also identified through citations.

B. Study selection process

The *Web of Science* search returned $k = 406$ articles (without duplicates). In the second step (Fig. 1, “Screening”), all of them were screened independently by two raters, according to inclusion criteria using the titles and abstracts of the articles (see Table I also for some

exclusion notes). Dissenting judgments occurred in less than 5% and were resolved through discussion. Unclear cases were marked as “review required,” and these studies were briefly skimmed for content in order to make a judgment. Full-text access was obtained for all remaining articles ($k = 77$), and the texts were reviewed again according to the inclusion criteria (Fig. 1, “Eligibility”).

In particular, a distinction was made between studies that are clearly related to PER and those that use only physics-related stimuli without focusing on physics education or physics learning. As we are interested in the specific conclusions and implications of the studies for physics education, for each article, an individual inclusion decision was made based on the theoretical background, the research questions or hypotheses the study answered, and the implications the article derived for learning. More specifically, we examined whether the studies drew on the PER literature, whether physics-specific questions were addressed, and whether physics-specific implications were given. The eligibility of 77 articles was discussed between both raters and the reasons for exclusion or inclusion were documented. Based on full-text review, it turned out that seven studies on reading comprehension were excluded despite using physics contexts as study material. More specifically, studies in this area compare the cognitive processing of refuting and nonrefuting texts [10,11] or use conspiracy texts [12], investigate word processing when replacing physical concepts with familiar words [13,14], investigate text comprehension as a prerequisite for science learning [15], or investigate the influence of mood on text processing and comprehension [16]. In these cases, the choice of physical examples were not crucial in terms of the research questions or implications of the studies. The same applies to the studies that primarily address multimedia learning environments [17–23]. In fact, all these studies are either covered by the review on eye tracking in exploring learning in general [6] or by the review paper on eye tracking in *Multimedia Learning* [7] if they were published within the time ranges covered by these reviews (i.e., by the end of 2012 or 2016, respectively). Further, three more studies were excluded because they deal with problem solving in general, focusing on gender differences [24], item difficulty and cognitive load [25], or gifted

TABLE I. Inclusion criteria and notes for exclusions.

Inclusion criteria	Notes for exclusions
The article has been published in a peer-reviewed journal in English.	The article is not written in English or it is not a journal publication ^a .
The article reports original and empirical research.	The article is a review or commentary article.
The article is in the field of physics education research.	The article uses physics-related visual stimuli, but other than that, had little to no relation to physics education or PER (as indicated by the educational background, research questions, and implications of the article).
Eye-tracking technology was used in the research.	

^aSources other than journal articles were documented to answer RQ1 and for the purpose of completeness.

students [26]. While all these pedagogical topics are peripherally related to PER, the argument of content replaceability also applies to these studies and, finally, to a study focusing on visual attention during instruction [27].

As described above, snowballing and reverse snowballing were used to identify cited and citing papers, including nonjournal articles (proceedings, book chapters, theses). These studies were put through the same procedure, yielding a total of $k = 56$ eligible records (Fig. 1, “Inclusion”; three proceedings had to be excluded because they were not available and could not be located through regional libraries) and $k = 4$ additional Ph.D. theses. Overlaps that occurred between different proceedings or between journal articles and other literature sources were documented. Thirty-three journal articles were included in our analysis for answering RQs 1–3, and an additional 23 resources (proceedings and book chapters) were considered with respect to RQ1 and the task type (visual stimuli; RQ2). Because of the large overlap between the Ph.D. theses and the associated journal articles, proceedings, and book chapters, the former were not considered for analysis regarding our research questions. Among the 33 journal articles, 27 were found by the *Web of Science* data search, and six articles were added by snowballing. The latter were published in rather “PER-atypical journals,” for example, *Applied Cognitive Psychology*.

It is worth noting that none of these 56 eligible studies have been reported by the reviews by Lai *et al.* [6] or Alemdag and Cagiltay [7], and only one study (thematically based around vector fields) is included in the review by Strohmeier *et al.* on eye tracking in mathematics education research [8]. This indicates minimal overlap of our review and the reviews conducted by others, thereby suggesting that the applications and implications of eye tracking that are specific for PER can be distilled from our selection of studies.

C. Coding scheme and analysis

The journal articles selected for this review were analyzed according to the research questions. A complete survey on all studies included in this review with selected information can be found in the tables in the Appendix. With respect to RQ1, we categorized (i) *publication* (year and journal), (ii) *eye-tracking scenario* (e.g., assessment), and (iii) the *topical area* (e.g., conceptual understanding) as well as the *physics domain* (e.g., kinematics; see Table VII). For the eye-tracking scenario, we were able to distinguish three learning scenarios in which participants’ eye movements were measured: First, if the eye movements were recorded while students were solving physical problems or working on tasks, then we grouped them into the *assessment* scenario. Typical task types include, for example, solving standardized assessment instruments such as the Test of Understanding Graphs in Kinematics [28] or the Force Concept Inventory [29], synthesis problems,

troubleshooting tasks, or the interpretation of representations. We also indicated whether the assessment took place once (single point), repeatedly (longitudinal), or as an adaptive testing sequence. A typical example of longitudinal assessment is the use of pre- and post-tests. Solving a series of tasks was not assessed as longitudinal (e.g., solving several questions of a concept test). If the test items were flexibly adapted during processing, we labeled this scenario as an adaptive assessment. For the second category, the participants acquired knowledge during an instructional design, referred to as the *knowledge construction* category. Examples include engaging with instructional designs, such as reading text, operating simulations, viewing animations, or teaching in the classroom. In the third category, students were involved into *laboratory work* including experimentation (virtual or real)—that is, operating equipment, collecting data, and analyzing data. Studies could also be assigned into more than one category (e.g., assessment and knowledge construction). The topical areas were classified based on the synthesis given by Dockter and Mestre of PER [30] and involved conceptual understanding, graph understanding, representations, etc. (cf. Sec. IV C for further details). The physics domain was directly mentioned in most articles or became obvious by examining the stimuli used.

With regard to RQ2, we extracted the methodological information from the studies, such as (iv) *stimuli* (task type and areas of interest), (v) *sample* [learners and sample; for (iv) and (v) see Table VIII], (vi) *device* (manufacturer, sampling rate, calibration, movement restrictions, distance between participant and screen, fixation detection algorithm, and accuracy information), and (vii) additional data (see Table IX). For the task type of the stimuli (iv) all 56 studies (journal articles, proceedings, and book chapters) were considered (cf. Sec. VA), although nonjournal articles are not presented in Table IX. For the journal articles, we considered the level of detail with which eye movements were investigated.

Further, we introduce the terms *global* and *local* to distinguish between analysis at the level of complete stimuli (e.g., a graph as a whole) and at a smaller level of visual units (e.g., axis labels, specific graph regions; see Fig. 2). More specifically, a *global* label was assigned whenever eye movements were examined based on obvious surface features of the stimuli—for example, the AOIs were defined to separately cover the stem and options of an item stimulus, or a learning page was divided into look zones that covered the introduction, main text, and diagram. If a stimulus was analyzed at the *local* level, the AOIs covered specific parts of the stimulus, such as relevant components of an electrical circuit or relevant areas within a graph (compared to viewing the diagram as one unit). By differentiating the analysis according to the level of granularity, we also echo the thinking of Andra *et al.* who distinguish a *macro-*, a *meso-*, and a *microlevel* of eye-tracking analysis based on the definition of AOIs [31]. In this regard, the

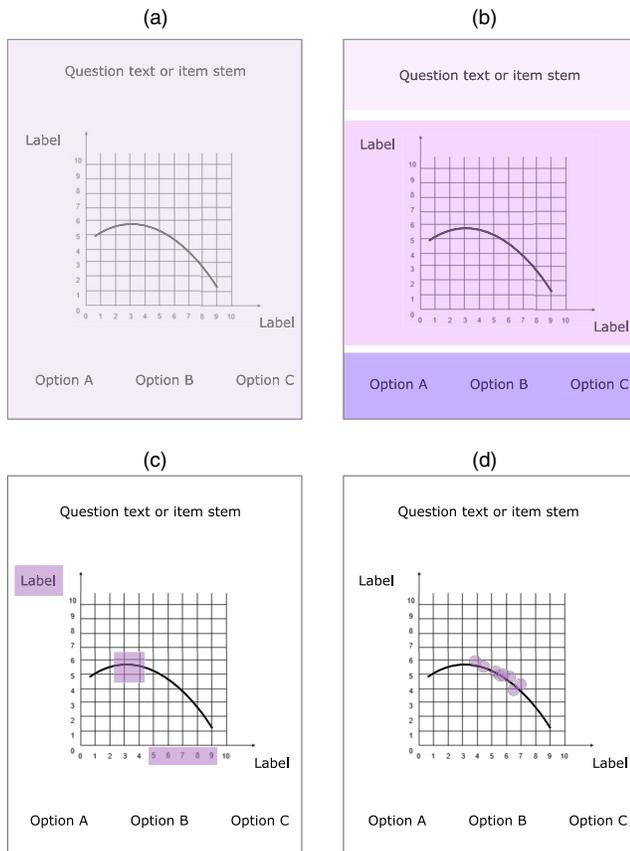


FIG. 2. Illustration of global and local analyses of a visual stimulus (RQ2). For (a)–(d), the figure shows the identical visual stimuli, comprising a question text, a diagram, and response options. We use the term *global analysis* when, for example, gaze behavior is considered on the whole stimulus (a). The subdivision into surface features of the task format (b) is also considered as a global analysis. If, on the other hand, task-relevant elements within a representation are considered, we use the term *local analysis* (c). This includes task-relevant or irrelevant areas, or areas that are of special interest for the research question under investigation. Also, when the temporal evolution of gaze behavior is considered within a representation, for example, whether eye movements follow the graph, we refer to this as a local analysis (d).

authors refer to a distinction from Holmqvist *et al.* that distinguishes *focused* versus *overview* eye-movement behavior on a stimulus to reflect different search and problem-solving processes [32]. To distinguish between local and global, it is important to look at how the eye-tracking data are analyzed in addition to the areas of interest (AOI) size. A local analysis involves examining the task-specific movements of the eye, for example, whether they depict the perceptual executive actions (as when tracking a graph to determine its slope). The analysis of expertlike or novicelike areas is also part of a local analysis. Global analyses aggregate the information content of eye movements and can often be performed equally for different stimuli, regardless of their content. Similar efforts can be found in the literature to subdivide this level of analysis

when a stimulus consists of multiple elements, for example, in problem-solving strategies versus navigation [33,34]. Independently of eye tracking, Brünken *et al.* refer to local and global coherence formation, referring to the processing of information within a representation and the linking between multiple ones, respectively [35]. In Rau's work this is reflected in the terms *visual understanding* and *connectional understanding* [36].

Last, subjects in the studies were classified as school students, university students (including college and undergraduate students and above), or mixed.

RQ3 led to the category (viii) *eye-tracking measures and variables* (metrics, theories, and split criteria for data analyses; see Table X). According to the categories defined above, we summarized the results of the studies, the framing theories for interpreting eye movements, and the implications and limitations. The relationship among the eye-tracking measurements, cognitive frameworks of interpreting eye movements, and learning performance was identified by synthesizing common and conflicting findings among the studies reviewed. Based on the reported findings, general conclusions were drawn with regard to how eye tracking provided insights into physics education and which limitations became apparent.

For all 33 studies, the information regarding RQ1–3 was extracted independently by both authors after agreement on a coding scheme and definitions of categories. The same procedure was applied for the 23 additional publications regarding RQ1 and the task type. Disagreements of double coding were discussed and resolved.

IV. DISSEMINATION OF EYE-TRACKING STUDIES IN PER (RESULTS PERTAINING TO RQ1)

We structure the results according to the research questions (see also the tables in the Appendix, Tables VII–X). For better readability, it must be noted that Ref. [1] as well as Refs. [37–68] refer to journal articles that are part of the review. Subsequently, the secondary sources [69–95] are listed for completeness.

A. Number of studies and publication information

For the journal articles, the earliest study was conducted in 2005 by Van Gog *et al.* [44], and it was found that an increasing number of studies have been published since 2014 (see Fig. 3). With 27 out of 33 journal studies published since 2014 (82%) and 22 out of 33 since 2017 (67%), the number of publications using eye tracking in PER has been growing steadily, while 2020 forms the present peak in the number of studies conducted, with seven studies published. The journals cover the fields of physics, education, psychology, and technology. In total, the publications were issued in 14 different journals, with the majority of the studies being published in Physical

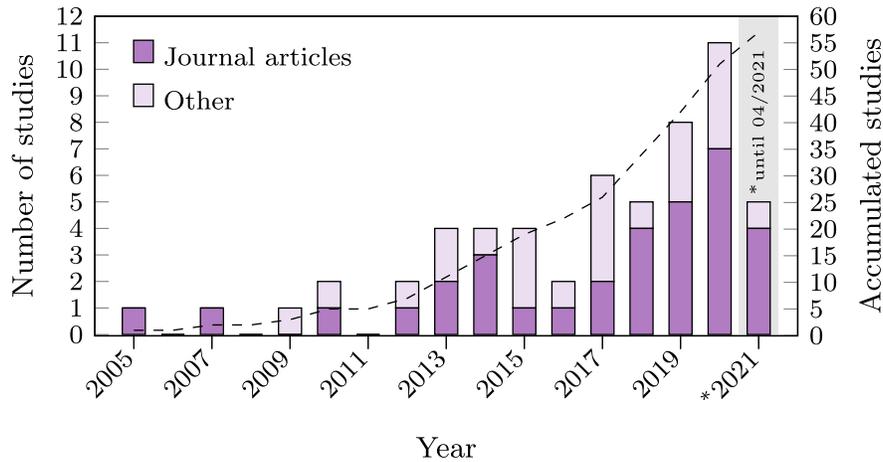


FIG. 3. Number of eye-tracking studies in PER per year (left vertical axis) and in total (right vertical axis). For 2021, all articles published until the end of April have been included.

Review Physics Education Research (16 studies or 48%). Two publications were published in European Journal of Physics, Computers and Education, Frontiers in Psychology, and Scientia in Educatione, respectively.

For the proceedings and book chapters, a similar chronology was found. From among 23 records, 18 were published since 2014 (78%) and 13 were published since 2017 (57%). The most common resource for eye-tracking studies in PER aside from journals are the PERC Conference Proceedings (5 studies or 22%), the AIP Conference Proceedings (5), and

the ETRA Conference Proceedings (4). Only three articles were published as book chapters.

The similarity in chronology to the journal articles is not surprising, as most work published in conference proceedings has been reconsidered for publication in regular journal articles (cf. Sec. IV C). Because of this canonical overlap, we always report results on both publication types separately to avoid aggregation bias of duplicate reports. A visual overview of these relationships between all proceedings and book chapters identified in the search

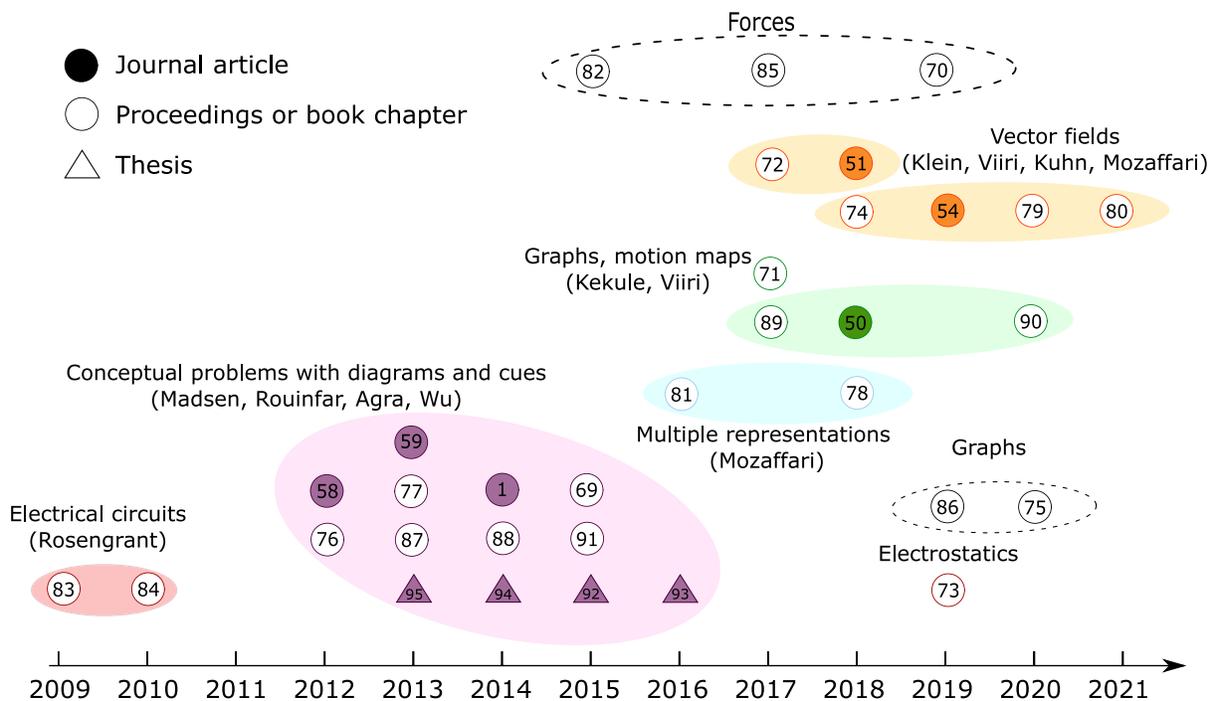


FIG. 4. A visual summary of all proceedings and book chapters identified in the search ($k = 23$; circles not filled in color). Associated journal articles (circles filled in color) and Ph.D. theses ($N = 4$; triangles) to these records are also presented. The colored ellipses indicate related studies from overlapping author groups. The dashed ellipses indicate thematically related studies from different author groups. The numbering refers to the references in the bibliography.

and the associated journal articles is presented in Fig. 4. For completion, the Ph.D. theses of the corresponding research groups, that were not considered for analysis, are also included here. The high degree of similarity between journal and nonjournal articles is also evident in the next sections and supports the decision of restricting further analysis to the journal articles.

B. Learning scenarios when eye tracking was used

For each study, we extracted the type of learning situation the participants were in while their eye movements were recorded. We distinguished between (i) *assessment* or *problem solving* (single point and longitudinal), (ii) *knowledge construction*, and (iii) *laboratory work*.

The categorization of the studies is documented in Table II (first column, italic categories). In order to clarify the

categories, we will show their distinction with some examples below. For example, the study by Han *et al.* [45] examined students' attention while solving the TUG-K at the beginning and the end of a physics course. This qualifies the study to be included in the (longitudinal) assessment category. Because eye-movement data was not collected during learning between the two tests, this study is not categorized as knowledge construction. In the study by Klein *et al.* [54], students were instructed via written texts on how to interpret vector fields. Before and after this intervention, students solved a series of tasks. In this case, eye movements were recorded during both (longitudinal) assessment and knowledge construction. In the studies by Rouinfar *et al.* and Madsen *et al.* students solved a series of conceptual tasks [1,59]. Based on their answers, students were given cues in follow-up tasks that visually highlighted conceptually relevant areas of the diagram. The authors also measured the

TABLE II. Dissemination and use scenarios of eye tracking. References to nonjournal articles are included in parentheses. Note that studies that appear in both the *assessment* and *knowledge construction* category used eye tracking in both scenarios.

Scenario and topical area	Physics domain	Visual stimuli or task type ^a
<i>Student assessment or problem solving</i>		
Conceptual understanding	Statics	CATS test items [37]
	Forces	FCI test items [45,56] ([70,82,85,90])
	Energy conservation, work	Items with diagrams [1,58,59] ([69,87,88,91])
	Various concepts	MC test items [41] ([76,77])
		Synthesis problems [48]
	Electrostatics	Integration of charge distribution ([73])
Graph understanding	Kinematics	Line graphs [38,39,52,64]
		TUG-K test items [49,53,55,56] ([71,75])
		Distance-time graphs [1,58,59]
		Vertical toss problem ([86])
	Various domains	Linear and nonlinear graphs [61] ([69,77,87,88,91])
Representations	Kinematics	Multiple representation tasks [68] ([78,81])
	Forces, motion, energy	Items with diagrams (R-FCI, other) [46,50,65] ([89,90])
	Electrical circuits	Troubleshooting tasks [44] Calculate resistance ([83,84])
	Wave optics	Interference and diffraction patterns [66,67]
	Vector fields	Interpretation of divergence or curl [51,54] ([72,74,79,80])
	Inertial forces	Trajectories in frames of reference [57]
	Coordinate systems	Transfer data to diagram [47,57]
	Measurement uncertainty	Data representations [63]
<i>Knowledge construction</i>		
Learning from text with representations	Forces (inclined plane)	Text and picture combination [60]
	Mechanics	Worked-out examples with text and math [62]
	Electricity	Online learning module [40]
	Vector fields	Instructional texts [54] ([74])
Computer-based learning	Kinematics, energy	Tasks with overlaid cues or hints [1,59] ([69,87,88,91])
	Optics (refraction of light)	Simulation interface [43]
	Coordinate systems	Multimedia instruction [47]
<i>Laboratory work</i>		
Experimentation	Ideal gas law	Conduct experiment, complete worksheet [42]

^a(Representational) Force Concept Inventory (R-FCI), Test of Understanding Graphs in Kinematics (TUG-K), Concept Assessment Tool for Statics (CATS), multiple choice (MC).

impact of these cues on knowledge acquisition during the testing period. We categorized these studies as adaptive assessment and knowledge construction, because, on the one hand, there was a test situation and, on the other hand, students' knowledge was specifically influenced by training.

Eye tracking was most frequently used in assessment scenarios, in journal articles (28 studies or 85%) as well as other publications (23 of 23), and independent from the publication type, longitudinal assessment was rare (6 of 28 and 3 of 23, respectively). Visual attention during knowledge construction was examined in eight journal articles and five other articles (24% and 22%), respectively. There is only one journal publication on visual attention during laboratory work.

C. Physics domains and topical areas

Among the physics topics, a large emphasis is placed on the domain of mechanics, such as kinematics, motion, (inertial) forces, energy, or work (21 of 33 journal articles, and 16 of 23 proceedings and book chapters). Other research interests included, for example, vector fields, electrostatics, measurement uncertainty, and (wave) optics. The coverage of the subject areas in the studies is documented in Table II (second column). Note that in order to obtain a better idea about the physics domains, the table also includes stimuli information (third column), although this belongs to RQ2(iv) and is reported in detail in the next section (cf. Sec. VA).

While reading the articles and examining the tasks for the participants, we noticed a few topical clusters that had accumulated. Some tasks, for example, required handling and working with discipline-specific visual representations (such as interference patterns or electrical circuits). Other problems were aimed at conceptual understanding. The learning scenarios often involved reading texts enriched with representations, dealing with multimedia learning environments, or testing instructional-design principles. This led us to assign the articles to topical areas in an inductive process—inspired by the work of Dockter and Mestre [30]—which are also listed in Table II (first column). Because of the cross-cutting nature of numerous problems, a few studies that were included in a particular section (e.g., conceptual understanding) could have also been included in another section (e.g., representations); therefore, this classification serves only to support an overview and we refrain from counting the number of studies that belong to individual topical areas. However, it is evident that there are emphases on graph understanding in kinematics and conceptual understanding of energy and work. Presenting the stimuli information used in the studies in Table II (third column) also provides a more complete picture of what has been investigated in the studies with respect to the topical areas.

Upon further examination of conference publications and book chapters, we found a large overlap with journal articles pertaining to the same subject area and physics

domain. In Fig. 4, all nonjournal articles (circles not filled in color) and the associated journal articles (circles filled in color) as well as theses (triangles) are shown as points on a timeline, and the thematic grouping is indicated by ellipses. For example, the studies reported in Refs. [74,79,80] use the same stimuli as in the journal publication [54]. Another overlap regarding research questions and stimuli also became apparent between the studies reported in Refs. [69,76,77,87,88,91] and the journal publications [1,58,59]; in this case, all studies were framed in four Ph.D. theses available online [92–95]. The illustration not only reveals the relationships among the studies but also the chronological progression of various projects.

V. METHODOLOGICAL IMPLEMENTATION AND VISUAL STIMULI (RESULTS PERTAINING TO RQ2)

A. Stimuli used in the studies and subdivision into areas of interest

The stimuli used can be inferred from Table II (third column). In the case of assessment scenarios, test items were used, which mostly consisted of a question statement (item stem), multiple-choice alternatives (options), and a diagram or graph. For the knowledge construction scenario, the stimuli consisted of learning materials—that is, written texts with illustrations presented on the computer or multimedia learning environments also viewed or controlled on the computer. In the study investigating laboratory work, students' visual attention on worksheets and equipment was measured.

To analyze the eye movements on the visual stimuli quantitatively, look zones, so-called areas of interest (AOIs), were defined in all studies with the exception of two [46,49].

The analysis of gaze behavior can be differentiated into two levels of finesse—that is, a global level and a local level. This distinction finds support in a critical discussion on the level of finesse used in the studies by Chien *et al.* [42], Küchemann *et al.* [57], and Klein *et al.* [55]. Smith *et al.* [62] also discuss *different levels of granularity* when they decompose their worked-out examples into text and equations, once globally for text and mathematics and once blockwise for each step of thought. Regarding the distinction in a global and a local level of AOIs [31], for example, Madsen *et al.* [58] used student interviews to identify novice- and expertlike AOIs in diagrams involved in conceptual physics problems regarding speed and energy conservation. These AOIs covered small areas within the diagrams, and students' visual attention to these areas was measured [1,58,59]. Thus, local eye movement evaluations can be used to investigate cognitive processes involved when using specific representations at a more detailed level compared to a global evaluation [48].

Table VIII in the Appendix presents the assignment to global and local evaluations for each study, specifying AOI

definitions. Overall, 27 studies subdivided the stimuli based on surface features (global), and 16 studies considered attention at a local level (11 studies conducted analyses at both the global and local levels). As mentioned before, two studies considered heat maps or gazeplots to qualitatively discuss the data and did not define AOIs at all [46,49].

B. Subjects

The participants in most studies were undergraduate students (17 of 33; including college students), followed by high school students (11 of 33; including junior, middle, and high school students). Mixed samples were used in five studies, including, for example, Ph.D. students, teachers, or faculty members. Three publications reported on multiple substudies [56,58,63]; in these cases, subject information was extracted only for the cohort that was examined with eye tracking. One study by Kozhevnikov *et al.* referred to students who were recruited from a general participant pool, without providing more detail [56]. However, the authors refer to a sample from another substudy, in which the sample is described to comprise undergraduate students.

In all publications, participants were tested individually. On average, each study included a sample of $N = 54.4$ participants ($SD = 29.9$), ranging from $N_{\min} = 15$ [56] to $N_{\max} = 119$ [47]. Seven studies reported exclusion of participant datasets for technical reasons [46,47,50,52,59–61]. In these cases, dropouts ranged from 2.9% to 32.4%, with an average of 16.1%. Four studies excluded participant data based on a performance criterion [42–44,59], like including only extreme performance groups.

C. Eye-tracking devices and calibration procedures

Eye-tracking devices can be divided into two classes: Remote eye trackers are statically attached to the stimuli, typically at the computer screen. In contrast, mobile systems are attached to the subject's head or are contained in glasses worn by the subject. In this review, only one study by Chien *et al.* used head-mounted eye-tracking systems [42], while all other studies used remote eye-tracking systems from leading manufacturers in accordance with Holmqvist *et al.* [3]—Tobii Technology (15 of 33), SensoMotoric Instruments (SMI; 8), and SR Research (6). It must be noted that studies from the same group of researchers are not independent of each other in terms of technological equipment. Therefore, these figures only provide an orientation and are not suitable for a quantitative comparison.

Further, the sampling rate of the eye trackers was mentioned in all but two publications. Depending on the device, it varied from 30 Hz for the eye tracker from the Danish company Eye Tribe to 1000 Hz for the EyeLink from the Canadian company SR Research. Most eye trackers recorded gaze data at a sampling rate of 120 Hz or below (18 of 31 studies that reported a sampling rate). The eye tracker's spatial accuracy was documented in approximately half of all studies (18 of 33) and was in all

cases reported as being below 0.5° . The average distance between the participants and the screen varied between 50 and 70 cm, as reported in 18 of 33 studies. The calibration procedure was mentioned in 22 studies and was reported as a 5-point (6 studies), 9-point (13 studies), or 13-point procedure (3 studies). Only three studies reported a threshold agreement regarding the calibration procedure [1,47,58], and another eight studies mentioned a verification of the calibration results [37,49–51,56,61,66,67]—for example, “calibration data and visual inspection of scan paths showed a reasonably good quality” [66] (p. 3). In addition, 21 of the 33 publications provided information on movement restrictions, where we judged statements like “the participants were asked to keep their head fixed” as head-free gaze capturing [63–65]. From among these 21 studies, seven reported using a chin or forehead rest (or both) [1,43,44,56,58,59,68]. In most cases, movement was restricted for the purpose of increasing measurement accuracy. In addition, 22 studies reported fixation identification algorithms, mostly referring to the classification of Salvucci and Goldberg [96]. In 14 studies, a velocity-threshold algorithm was used to separate and label fixations and saccades, five studies used a dispersion-based identification, and three studies reported using an algorithm that was not further specified. Moreover, all studies were examined regarding a specification of the captured eye positions as an indicator for the accuracy and validity of the data obtained. However, since this information was only reported in one article (stating that “the eye positions [were recorded] during approximately 90% of time” [66] (p. 3)), this aspect was not included in Table IX in the Appendix.

D. Additional data

Apart from gaze-data analysis, approximately two-thirds of the studies collected additional data from the subjects as control variables. Most frequently, oral or written explanations from the students, spatial skills, confidence or mental effort ratings are collected to better interpret the eye-tracking data. Often the additional data are used as split criteria for the analyses of the eye-tracking data; in the following sections, these analyses and the independent variables against which the eye-tracking metrics were evaluated are examined (cf. Secs. VI and VII).

VI. DEPENDENT AND INDEPENDENT VARIABLES (PREPARING THE RESULTS PERTAINING TO RQ3)

A major challenge in eye-tracking data collection is establishing a link between the collected data and the underlying cognitive processes. Here, the selected eye-tracking metrics and their theoretical interpretation play a crucial role. In this section, we report the metrics and variables that were used, including the dependent and independent variables that were related to the eye-movement

measures. By doing so, we prepare the structure of the results and implications obtained in the studies (Sec. VII).

Two studies qualitatively interpreted eye movements exclusively using gazeplots or heat maps [46,49]. The remaining 31 studies reported at least one of the following metrics, with the number in parentheses denoting the frequency: Total fixation duration or viewing time (also referred to as dwell time) (25), fixation count (16), mean fixation duration (8), transition count (8), saccadic angle (4), saccade length (3), and other (such as rereading time or time to first fixation). If studies reported metrics as ratios or proportions, the associated primary measure was noted—for example, total fixation duration per area was scored as (total) fixation duration. An overview and definition of the most common eye-tracking metrics is provided in Table III.

In addition, all independent variables were extracted for which the eye-tracking data have been evaluated referring to qualitative as well as quantitative analysis approaches. Independent variables can either be defined by the study design (as an *a priori criterion*) or they can result from one or more *post hoc* splits of the datasets (e.g., comparing visual attention between correct and incorrect responses). We identified the most frequent independent variables in the following manner and summarize them in Table IV.

- (i) *Performance*.—In most of the studies (23 of 33), the eye-tracking metrics have been compared between correct and incorrect responses, between high and low confidence responses, or between overall very good and very poor learners (low vs high performers or experts vs novices). Here, the

performance assessment was not based on an external criterion but inferred from test scores obtained during eye tracking.

- (ii) *Stimuli*.—In 19 studies, the stimulus was varied and the influence of the different stimuli on eye movements was investigated. Typically, stimuli contrasts are used in studies comparing learning with different materials—for example, quiz vs homework problems [62], text with or without representation [40,60,63], or animation vs interactive training [47]. This criterion also applies when the participants worked on different types of tasks—for example, Brückner *et al.* [39], Susac *et al.* [64], and Klein *et al.* [52] compared eye movements for different graph task types (qualitative vs quantitative or slope vs area concept).
- (iii) *Progression*.—In 10 studies, visual attention was compared among different time points or phases. This applies, for example, when a stimulus is presented repeatedly (pre- vs post-test [39,45,54] or initial vs transfer problem [1]) or when there are different experimental or problem-solving phases [42–44,57]. This category also applies if the gaze data is divided, *post hoc*, into small time periods to look at progression of visual attention over the time course of reading or problem solving [41,43,54,59].
- (iv) *Students*.—In eight studies, visual attention was compared between different groups of students (here, group membership is not defined by using different materials or stimuli [see (ii)]). A few

TABLE III. Overview of the most common eye-tracking metrics reported in the PER studies and their relevance.

Metric	Definition and synonyms	Use and interpretation
Fixation count	The number of fixations	Often used as a measure of attention directed to defined areas of visual stimuli. High numbers indicate more attention.
Total fixation duration	Dwell time, total visit duration, or viewing time; sum of all fixation durations and saccade durations	Similar to fixation count, (total) fixation duration measures how long information is accessed. Higher fixation durations indicate more visual attention to certain areas.
Mean fixation duration	The average duration of single eye fixations typically range from 100 to 600 ms.	This metric is often considered an indicator of cognitive processing demand and may also indicate, for example, the effort required to discard individual options in multiple-choice assessment scenarios. Higher values typically indicate a higher cognitive effort to process information.
Transition count	Number of gaze shifts from one area to another area	Indicates integration processes among different sources of information—that is, for example, to study gaze switches between options and question stem in multiple-choice scenarios or to study text-image integration during reading.
Saccadic length	Spatial distance between two consecutive fixations	Can be used to explore reading behavior.
Saccadic angle	The angle between the horizontal and the saccade	Can be used to investigate the direction in which learners move their eyes over visual representations. In graph tasks, for example, saccade angle information can be used to investigate whether learners trace the graphs with their eyes.

TABLE IV. Structuring the study results according to the split criteria used to analyze eye movements.

Independent variable	Contrast	Specification (studies)
Performance	Response accuracy (correct vs incorrect, option choice)	Impact on fixation measures on global and local levels [1,37,38,41,43,45,49–51,53,53,58,62,65–67], transition counts [38,43,44,48,55], and other (gaze) metrics (e.g., saccadic length) [41,43,51,54]
	Achievement (high vs low)	Impact on several indicators of visual attention [40,42,44–46,49–51,53,54]
	Confidence (high vs low)	Impact on several indicators of visual attention [53,55,57]
Stimuli	Representations (task type, strategies)	Investigating graphs in kinematics [39,52,55,64], multiple representations (motion maps, kinematics, and vector fields) [48,50,51,68], or texts and diagrams [41,63,65]
	Learning materials	Text-picture combinations [40,47,60,62], visual cuing [1,54,59], laboratory work [42]
Progression	Repeated presentation (e.g., pre- vs post-test)	[1,39,45,54]
	Separating different phases of learning	[42–44,57]
	Exploration of learning progression	[41,54,59]
Students	Spatial ability (high vs low)	[56]
	Expertise	[45,61,67,68]
	Study background (physics vs nonphysics)	[39,52,64]

studies employed an external performance criterion to separate student groups [56,61,68], whereas other studies examined students with different study domains (e.g., physics vs psychology students) [39,52,64] or students who attended different courses prior to the assessment [67]. In the study by Han *et al.* [45], a pre- and post-test design with two randomly selected groups from the same (on-going) physics course were used with several weeks between pre- and post-test. In this case, the student and the progression criteria apply in the same manner.

It is not surprising that certain studies use multiple split criteria to answer the research questions and, therefore, also consider interaction effects—for example, between stimuli and students [39,52,64,68]. Moreover, certain studies compared students' visual attention on different AOIs for the same stimuli (e.g., globally, that is, the time spent on the options vs the time spent on the item stem of a question). Information on the segmentation of stimuli into AOIs has already been provided in Sec. VA, and we did not use the AOI as a split criterion or as the independent variable.

The next section is divided according to the factors identified here and thoroughly refers to the eye-tracking scenarios identified in Sec. IV B (*assessment, knowledge construction, and laboratory work*). Whenever possible, sections are also meaningfully sorted by the eye-tracking metrics that represented the target variables in the studies.

VII. FINDINGS, INTERPRETATIONS, AND LIMITATIONS (RESULTS PERTAINING TO RQ3)

A. Performance-based analysis of eye-movement data

In assessment scenarios, several studies report a link between visual attention and the accuracy of the given solution or the expertise of the students. Visual attention has also been compared between successful and less successful students in learning contexts. Here, we summarize the main findings of the studies with respect to performance measures. An overview of the results presented here can be found in Table V.

1. Linking eye-tracking indicators with response accuracy

Total viewing time.—Almost uniformly, the total time spent on the tasks was reported not to correlate with accuracy, either in (kinematics) graph tasks [53] or in conceptual tasks related to statics [37] or forces [45]. Susac *et al.* also reported similar dwell times on questions between student groups despite differences in performance on the wave optics test [67]. In addition, Klein *et al.* as well as Smith *et al.* found no relationship between time spent on instruction and student performance on the target problems on vector fields and mechanics, respectively [54,62]. Finally, Susac *et al.* reported eye-tracking data and response accuracies for solving six questions on energy (with and without diagrams) [65]. Since only in one task a correlation between response accuracy and eye-tracking data was tested, we disregard the study at this point.

TABLE V. Overview of performance-related results from the eye-tracking studies. The results are phrased as statements, and studies that do or do not support the statements are listed.

Statement	Supporting studies with topic	Rejecting studies with topic
Total viewing times or fixation counts on the stimuli are not correlated with response accuracy	<i>Assessment:</i> Kinematics graphs [53] Statics [37] Forces [45] Wave optics [67] <i>Knowledge construction:</i> Vector fields [54] Mechanics [62]	
Longer viewing times or higher fixation counts on correct options or relevant parts of the stimuli are associated with correct answers	Wave optics [66] Graphs in kinematics [38,49,53] Motion maps [50] Forces [45] Vector fields [51] Energy and motion [1,58] Light refraction [43]	Statics [37]
Higher transition counts are associated with better performance	<i>Knowledge construction:</i> Light refraction [43] Electrical circuit [40]	Assessment scenarios: Synthesis problems [48] Troubleshooting tasks [44] Graph tasks [38,55]
Longer mean fixation durations and shorter saccade lengths are associated with better performance	Conceptual questions [41]	Vector fields [51] Energy questions with diagrams [65]
High-performing students devote their attention more quickly to relevant parts of the stimuli, have longer viewing times or higher fixation counts on relevant parts of the stimuli and use different strategies	Graph tasks [46,49,50] Troubleshooting tasks [44] Laboratory work [42] Vector fields [51,54]	(Graphs in kinematics [53]) (Forces [45])
Shorter total viewing times or lower fixation counts on the stimuli are associated with higher confidence	Kinematics graphs [53,55,57] (Wave optics [67])	...

However, if one applies a more detailed analysis—for example, by examining the tasks individually or dividing them into different areas (on a global or local level), the differences in visual attention become visible, as is shown in the next paragraphs.

Time spent on different parts of the stimuli.—Considering multiple-choice test items, students who answered correctly were found to have spent more time on the correct option than on others and, conversely, students who answered incorrectly spent more time on incorrect options. This was found for the topics of interference patterns in wave optics [66], graphs in kinematics [38,49,53], motion maps [50], and forces [45], but could not be confirmed for conceptual questions on statics [37]. For statics questions, it was found that correct solvers (i.e., those who provide accurate responses) fixate less on the questions and accompanying diagrams since they can extract key information faster [37]. In other contexts,

correct responses were also related with shorter times on globally defined parts of the stimuli—for example, when correctly judging the divergence, students tend to spend less time looking at the vector field plots [51], at least when using one of the two strategies investigated by Klein *et al.* Furthermore, in wave optics it was found that students attend to most of the patterns that they identify correctly [67].

When learning to conduct an optics simulation, Chiou *et al.* found that students with correct parameter choices took a shorter time to read the guiding question and set up the simulation than students who made incorrect choices [43].

Time spent on relevant and irrelevant areas in visual representations.—There are consistent results that indicate that correct solvers look more at relevant areas—for example, in kinematics graph tasks [38] or conceptual questions on energy and motion (in most cases) [1,58]. Madsen *et al.* found that students who answered incorrectly spent a higher percentage of viewing time looking at areas

of the diagram consistent with novicelike responses that represented common misconceptions that are well documented in the PER literature, whereas students who answered correctly spent more time on thematically relevant AOIs [58]. This finding was confirmed in the follow-up studies by Rouinfar *et al.* [1]. In addition, Chiou *et al.* found that students who failed to set up an optics simulation gave less attention to relevant parts of the simulation [43].

Transition counts.—To investigate the relationship between gaze switches and response accuracy, Ibrahim and Ding and Van Gog *et al.* examined transitions between different parts of the diagram in synthesis problems [48] and between different parts of electrical circuits in troubleshooting tasks [44], respectively. The researchers found no correlation between accuracy and the number of gaze transitions. Moreover, Klein *et al.* found that gaze transitions between options and questions or between different options were not correlated with performance in kinematics graph tasks [55]. In a similar context, Brückner *et al.* also confirmed that an isolated analysis of fixation durations and transitions provided only limited insight into graph-solving behavior. However, by applying a more detailed analysis combining fixation durations and transition metrics, Brückner *et al.* were able to show that correct solvers verify solutions more often by cross-checking different options and relating them with the item stem, thereby indicating that a relationship between accuracy and gaze switches cannot be precluded *per se* [38].

In scenarios of knowledge construction, two studies successfully related gaze transitions with response accuracy. Chiou *et al.* reported that successful students performed visual transitions among relevant areas of the simulation to observe the effect of parameter variations on light refraction [43]. Chen and She found that high-performing students put in more effort to assign correspondences between a water circuit and an electrical circuit when reading metaphorical texts on parallel circuits [40].

Other gaze metrics.—Mixed results were reported with respect to mean fixation duration and saccade length. Chen *et al.* found that correct answers to conceptual physics questions were associated with longer mean fixation durations (for question statements and diagrams) and shorter saccade lengths (for pictures) [41], whereas no such associations were found in the context of vector fields [51] or energy questions with diagrams (considering mean fixation duration) [65]. Further, the correct use of partial vector derivatives in the context of vector fields was reflected in the distribution of saccade angles. The saccadic angles reliably indicated whether or not students followed changes in field vectors in the appropriate direction or not [51,54]. This result became clearer when a stronger contrast was made between groups by comparing high-performing vs low-performing students, as is shown in the next section.

2. Correlation between overall performance and visual attention

In addition to the difference between correct and incorrect answers, certain studies also examined differences between high- and low-performing participants, thereby revealing a stronger contrast among subjects. Here, we summarize the results obtained from a relative distinction between achievement groups from the assessment (*post hoc*). At the global level, low achievers were found to spend more time on item questions and options (e.g., on graph tasks [46]), show less orientation through trial-and-error approaches (e.g., on electrical circuit debugging [44]), and browse through misconceptions by estimating them option by option [46,50]. In addition, for real laboratory work it was shown that low achievers needed more time to perform experiments and re-did them more often than high achievers [42]. Moreover, low-performing students also experienced high cognitive load in planning, conducting, and improving the experiments, as well as in interpreting the results, as indicated by high mean fixation durations [42].

In contrast, high-ability learners eliminate incorrect solutions more quickly by comparing relevant options and identifying differences and key features [46,50]. Consistent with the results reported above, high-achieving students spent more time on relevant areas of the problem and focused their attention more quickly on relevant sources [46]—for example, they focused on critical components in electrical circuits [44]. A deeper analysis of eye-movement data revealed that high-performing students also used other strategies during problem solving. They were found to focus their attention on specific areas in the field plot when interpreting vector fields and to exhibit sharp distributions of saccade angles in the horizontal and vertical directions [51,54]. In kinematics, they considered the graph as a whole rather than focusing on smaller regions [50], and in electrical circuit debugging, they spent relatively more time before deciding on a direction and evaluation [44]. When comparing mean fixation durations between low- and high-achieving students, Van Gog *et al.* found that high-achieving students have relatively shorter mean fixation durations in the problem-orientation phase and higher mean fixation durations in the problem-solving phase [44]. This is consistent with the finding of Klein *et al.*, which demonstrated higher mean fixation durations for high-performing students during the interpretation of vector field plots; these students also had shorter saccades when comparing adjacent vectors [51]. Despite these differences between high- and low-performing students, Klein *et al.* and Han *et al.* reported that high-performing students still devoted some of their attention to popular incorrect choices, thereby indicating conceptual mixing despite accurate responses [45,53].

3. Response confidence and visual attention

In three studies, a relationship was established between visual attention and the level of response confidence. Low confidence was consistently reported to be related to longer visit durations (for both correct and incorrect responses) [53,57]. Students with low confidence may consider more options, compare them, and take more time to select an option. Analyzing the gaze transitions between questions and options of TUG-K items, Klein *et al.* reported a difference between high and low confidence responses, which was much larger than between correct and incorrect responses [55]. This result suggests that gaze transitions may be more strongly correlated with confidence than accuracy at the global level. Other studies that collected confidence ratings did not relate them with measures of visual attention [39,52,54].

For tasks related to wave optics, Susac *et al.* also assumed that shorter means for dwell times were associated with higher response confidence, which put into perspective the elevation of confidence ratings [67].

B. Stimuli-based analysis of eye-movement data

We separately report the results related to the comparison of different task types and the influence of different learning materials on gaze behavior. The stimuli-based analysis also elucidates a few properties regarding the representations themselves that are mentioned here, for example, with regard to visual attention when working with graphs.

1. Processing graphs in kinematics

In the context of kinematics graphs, Susac *et al.* reported that students spent more time on diagrams when evaluating the area under the curve compared to the slope of the (line) graph [64]. Moreover, students spent similar amounts of time on the graphs when performing qualitative and quantitative evaluations; however, students fixated on the graph more often when the area had to be evaluated qualitatively compared to the slope. The follow-up studies by Klein *et al.* and Brückner *et al.* confirmed these results using the same test items [39,52]. For qualitative slope questions, further data analysis of saccadic directions has revealed that students performed eye movements corresponding to the gradient angle of the graph; hence, the eyes followed the graph, thereby suggesting that slope is an intuitive idea [52]. Longer viewing times on quantitative slope questions as compared to qualitative questions can be attributed to longer viewing times on the axes, as more information has to be extracted in the former [39,52,64]. Here, eye-tracking data can provide evidence for point-interval confusion or slope-height confusion when specific data points are fixated (or not) [52]. Moreover, eye-movement data revealed that students made an effort to understand the axis labels and, in the case of unfamiliar axis labels (e.g., changing the context from kinematics

to finance), the labels received more attention [52]. Using a dataset of the Test of Understanding Graphs in Kinematics (TUG-K), it was also revealed that the postulated requirements of the graph tasks are reflected by the eye movements when solving the tasks [55]. For this purpose, the gaze transitions between the individual elements of the questions (item stem and answer options) were extracted and subjected to a cluster analysis. The clusters corresponded exactly to the postulated task groups, which required a mapping from graph to graph, text to graph, and vice versa.

2. Revealing different visual strategies for interpreting (multiple) representations

The differences between multiple-representation tasks were addressed in a study by Kekule and Viiri in the context of kinematics [50]. They found that expertise showed up differently for the two types of tasks—that is, for graphs by looking at them as broadly as possible and for motion maps by looking at individual points in more local detail and comparing them with each other.

Wu and Liu investigated how students coordinate their attention when presented with multiple representations of a motion process [68]. Students were asked to evaluate the accuracy of statements regarding the movement based on the representations and to generate conclusions from the data. The researchers found that the different task demands were associated with different gaze transitions between the representations.

Further, Klein *et al.* reported that different visual strategies for interpreting the divergence of vector fields induced different visual processing of field diagrams, thereby reflecting different cognitive demands [51]. In an integration approach, students performed more fixations of longer duration and had shorter saccades compared to a differentiation approach. In the latter, eye movements reflected skimming the field and looking at adjacent arrows.

Extending previous research on single-concept problems, Ibrahim and Ding found that different kinds of synthesis problems were processed differently by students [48]. For simultaneous problems, students performed less within-diagram transitions and integrated visual information from the diagram and the text to make sense of the situation. In contrast, sequential problems evoked more within-diagram transitions and, thus, students made greater cognitive attempts to process the diagrams and treated each diagram more as a single event. Therefore, eye tracking reflects the different demands of both problem types in terms of linking multiple events and concepts.

3. Adding diagrams to test items

Apart from comparing different item types, several studies also investigated students' visual attention on tasks with or without accompanying pictures or diagrams. Susac *et al.* found that adding graphical data representations to data tables did not change the total time spent on

measurement uncertainty tasks; instead, it changed the distribution of visual attention [63]. With graphical representations, students looked less at the numbers or data tables. Thus, graphical representations of measurement data helped students to visualize the data presented, to focus attention on important features, and to reduce cognitive load by providing more resources for processing the data and making sense of it [63]. In a follow-up investigation using conceptual energy items, Susac *et al.* confirmed that total viewing time remained constant between both conditions (with or without supportive diagram) [65]. A more detailed analysis revealed that dwell times and number of fixations on the item stem decreased when a diagram was presented, thereby indicating that students spent less time reading the text of the problem and providing faster orientation for the initial phase [65]. Chen *et al.* also found that physics concepts can be represented more efficiently by pictures than by text [41]. Because of the sequential processing of text, areas that are crucial can be identified more quickly in images.

4. Text-picture combinations

Chen and She compared learning about electrical circuits between metaphors and analogies using texts and text-image combinations [40]. Learning with analogies resulted in stronger and longer integration processes compared to metaphors because they emphasize a one-to-one correspondence between the electrical circuit and the water pipes. In contrast, texts with metaphors required more attention overall to look for corresponding parts between systems because the structural relationships were more implicitly given. With regard to the presentation modality, learning with pictures evoked more integration processes between systems overall than learning with texts [40]. Accordingly, the pictorial elements can better lead learners to extract and infer relevant information, thereby leading to better understanding. Transitions were also indicative of learning in the study on coordinate systems by Hoyer and Girwidz [47]. They found that deeper processing of information occurred with interactive learning materials since students performed longer saccades corresponding to transitions between the important components of the learning material. Similarly, Smith *et al.* reported that students attended to conceptual information presented by text in worked-out mechanics problems [62]. The students also performed numerous transitions between text and math, thereby indicating learning and integration of information, independent of using the examples in a quiz or homework condition. Against the authors' expectations, the students incorporated textual explanations as an integral part of their approach when studying worked-out examples. Consistent with these results, Mason *et al.* found that combining an abstract illustration of an inclined plane with text promoted more attempts to integrate information, thereby resulting in better performance compared to using

concrete illustrations [60]. The abstract material promotes more efficient processing, as evidenced by shorter time spent on the text and shorter rereading time.

5. Visual cueing

Three studies have compared the effect of visual cues on students' attention. Madsen *et al.* found that participants who were shown cues after giving incorrect answers to conceptual questions localized more attention in expertlike AOIs and avoided novicelike areas in a transfer problem [59]. Consequently, seeing cues influenced their visual behavior, which can be interpreted as the first step towards thinking correctly about the problem. Consistent with this finding, Rouinfar *et al.* reported, in a follow-up study, that students who saw a cue were more likely to switch from an incorrect to a correct response [1]. In the cued condition, more participants displayed learning behavior after several trials of training with cues. However, students who changed their answer from incorrect to correct spent less time on the expertlike AOIs in the transfer problem, thereby indicating greater automaticity in extracting relevant information. This was also supported by the lower average fixation time on the expertlike AOIs. Thus, students can automate the extraction of relevant information through cuing [1]. Klein *et al.* used two divergence instructions with written explanations, equations, and visual representations that differed in the presence of visual cues within the diagram [54]. Students' eye movements were recorded as they processed the written instructions. It was found that the group with cues paid more attention to the diagram, made more transitions between the text and the diagram, and achieved better learning outcomes than the group without cues.

6. Laboratory work

Chien *et al.* were able to demonstrate that virtual and real laboratory scenarios are treated with different learning strategies using eye tracking [42]. In the virtual lab, students found it easier to focus on task-relevant zones in the experiment, whereas the worksheet received less attention. Consequently, they conducted more experiments than in the hands-on lab group. There was deeper cognitive processing of the questions on the worksheet; students thought and planned before they did anything.

C. Progression-based analysis of eye movements

1. Comparing visual attention between the pre- and post-test

Brückner *et al.* recorded students' eye movements while responding to line graph tasks at both the beginning and end of a lecture period [39]. They found that the results reported above regarding task type (qualitative vs quantitative) and concept (slope vs area) remained stable over time, whereas overall viewing times decreased from pre- to post-test. Specifically, dwell time on the options remained

stable from pre- to post-test, while dwell time on the questions and graph decreased. The researchers concluded that there were no recall effects regarding task solutions. In addition, no differences were found in dwell times between familiar and unfamiliar axis labels [39].

Further, Han *et al.* compared visual attention during problem solving at the beginning and end of a lecture period using the FCI in a web-based assessment [45]. While students' conceptual understanding increased significantly from the pre- to the post-test, the total time spent solving the FCI did not change; however, it must be noted that no matched samples were used. By analyzing the time spent on the different options of the FCI questions, they found that students' visual attention followed the conceptual progression in such a manner that students shifted their attention more to the expert choices in the post-test (while still maintaining a high level of attention on the naive choices, as reported above).

In contrast to these two studies, which reported differences on a fairly large timescale (several weeks between the pre- and post-test), four studies reported immediate effects of brief interventions on students' gaze behavior. When students were required to judge the divergence of vector fields in an initial task prior to any intervention, Klein *et al.* found that students focused primarily on the center of the vector field plots [54]. This area was perceptually salient and visually attractive because the direction of the arrows changed. Students were found to make eye movements in many directions, thereby indicating a nonsystematic sweeping of the field. Thus, students' visual attention was not guided by conceptual thoughts or heuristics. After the intervention, students performed predominantly horizontal and vertical saccades, and the distribution of saccade angles had peaks in these directions. This result reveals that students' gaze behavior changed meaningfully as a result of the intervention. As reported above, Rouinfar *et al.* also studied how students' attention changed as a result of repetitive training with or without cues [1]. The cued students' showed greater automaticity when extracting information from relevant areas of the diagrams, as reflected by less time spent on the areas. In line with these results, Küchemann *et al.* reported that the total viewing time for identical test items decreased when comparing the predict and observe phases before and after students had viewed a demonstration experiment on inertial forces [57].

2. Separating different phases during learning or problem solving

Visual attention data was used to separate different phases in the learning process; for example, when students explored a simulation on the refraction of light [43], worked through a worksheet for conducting experiments [42], or troubleshooted electrical circuits [44]. By breaking down the learning process into a setting-up and

implementation phase and analyzing visual transitional patterns within the phases, Chiou *et al.* concluded that the sequence of visual transitions between AOIs is an important factor that determines the outcome of learning with simulations [43]. Based on this result, they advocated using step-by-step guidance for making adequate visual transitions in order to observe and comprehend the target phenomena. Further, Van Gog *et al.* found differences in attention allocation between the problem-orientation phase compared to the problem-solving phase, thereby reflecting the perceptual encoding processes and careful examination, respectively [44]. Chien *et al.* sampled three sets of eye-tracking data, each encompassing crucial components in scientific inquiry, including messing about, reading, planning, and conducting experiments, generating research questions, and designing new experiments [42]. By doing so, the researchers identified what elements of the worksheet or experiment were fixated upon with what frequency for the different phases during experimentation, and they identified different learning behaviors between simulation-based experiments and microbased laboratories.

3. Exploration of learning progression

In order to analyze processing of the instructional page, Klein *et al.* divided the reading time into several time intervals of equal width for each student [54]. Within each time interval, the number of fixations on each AOI and the transitions between the AOIs were determined and then compared between the stimulus conditions (instruction with vs without cues). The time series (also referred to as gaze probability analysis) showed that cognitive integration processes occurred simultaneously in both groups, but were significantly more pronounced for students using cues, thereby reflecting a greater attempt to construct a coherent mental representation during the learning process. Chen *et al.* conducted a deeper examination of the order of fixation points as students responded to conceptual physics problems [41]. They found that the fifth and ninth to eleventh fixation points, in terms of their position and average duration, were critical in predicting the accuracy of the student's response. In a similar experimental setup to the one by Rouinfar *et al.* described above [1], Madsen *et al.* found that the scan paths of correctly answering students were similar for repeated trials, thereby indicating that similar gaze behavior is established over time for comparable tasks [59].

D. Students

1. Comparing visual attention between high- and low-spatial ability students

Bi and Reid, Hoyer and Girwidz, Mason *et al.*, and Klein *et al.* state that spatial reasoning ability played an important role in solving the physics problems in their studies and additionally examined this variable [37,47,54,60].

However, they did not relate eye movements to spatial ability, but used the variable to rule out *a priori* differences between groups or to establish the relationship between spatial skills and performance.

On the other hand, Kozhevnikov *et al.* used a paper folding test to estimate students' spatial ability and then defined low- and high-spatial ability groups [56]. The students' eye movements were recorded while the students solved an extrapolation problem (i.e., a moving puck receives a kick and the resulting motion is to be described) and a graphing problem. The researchers found that students with high spatial abilities imagined the path the puck would travel, as indicated by fixations in that region. Thus, they directed their attention away from the elements depicted in the figure to empty areas. These students also visualized the results of the two motion components by upward and rightward saccades. For the graph problem, Kozhevnikov *et al.* found that students with high spatial awareness were better at translating graph information into integrated functional relationships.

2. Comparing visual attention among students with different levels of expertise

A few studies used an external criterion to define expertise and compared visual attention during problem solving for the different expertise groups. Han *et al.* used the FCI in two groups of students [45]. One group was selected at the beginning of the semester, while the other group was selected towards the end of the semester. Since this is also a progression-based analysis, the results are already summarized in Sec. VII C 1. Susac *et al.* investigated the understanding of interference patterns between two groups of students who had previously taken different courses [67]. The experimental group performed hands-on experiments on the topic, while the control group experienced lecture demonstrations. They found that the experimental group spent less time on the choices of interference patterns and revisited them less often before responding. This suggests that the students who performed hands-on experiments may be more familiar with interference and diffraction patterns and were more confident in selecting the correct pattern [67]. In the study of Škrabánková and Beitlová, the expertise groups were divided based on teachers' judgments prior to the test [61]. This turned out to be problematic because the division did not correspond to the subjects' performance and visual attention also did not indicate adequate group division.

Further, Wu and Liu also used an external criterion to define expertise—that is, the number of points achieved in TUG-K [68]. The students were then divided into high and low achievers and worked on a problem with multiple representations in kinematics. In this manner, the researchers were able to investigate which representation (table, equation, or graph) was focused on more often by which group and how often transitions took place. Wu and

Liu reported that the group with a high level of prior knowledge made more transitions overall, more often fixated the equation, and—depending on the type of task (identifying content, establishing connections, or making predictions)—focused either on tables and equations or graphs more often. The group with low prior knowledge mainly made transitions between the question and the table, while the group with high prior knowledge made transitions between question and equation or between table and graph. Thus, high prior knowledge students were found to be more inclined to infer and integrate information among multiple representations, skilled at flexibly moving across them and connecting or transforming representations [68].

3. Comparing visual attention between physics and nonphysics students

Susac *et al.* compared physics and nonphysics (psychology) students' understanding of line graphs in different contexts while capturing eye movements [64]. Klein *et al.* and Brückner *et al.* reported follow-up studies using identical materials and a different sample of nonphysics (economics) students [39,52]. In these studies, participants can also be viewed as discipline experts or novices (with high or low prior knowledge on average) located at different points on the knowledge continuum. All studies reported that physics students spent longer periods of time dealing with problems they were unfamiliar with (i.e., financial graph tasks) than with known problems (i.e., kinematics graphs). Susac *et al.* found that novice learners (i.e., the psychology students) spent less time on problems that they could not solve and that were very difficult for them, which represents novel findings that are in contrast to the results obtained in other studies comparing expertise groups [64]. However, this was not confirmed by the follow-up studies on economics students. Instead, it was revealed that the economics students had similar viewing times on the diagrams as physics students, but their visual attention distribution was more novicelike [39,52] (see Sec. VII A for more details).

E. Theoretical framing

A particular challenge in the use of eye tracking is the meaningful interpretation of the gaze data. Because of the great freedom in terms of evaluation possibilities, which can lead to different results depending on the choice of metrics and the definition of the AOIs, a theoretical basis is important in order to enable a well-founded interpretation of the data. Numerous theories of cognitive psychology provide reference points for this. Less than half of the articles (14 of 33) explicitly reported a (cognitive) theoretical anchoring of the eye-tracking method. Almost all of the 14 articles that theoretically embed their analyses referred to Just and Carpenter's *eye-mind assumption* [97] (13 of 14 articles), which states that a fixation point of the eye corresponds to a fixation point of attention. Thus, gaze

duration is an indicator of the time required to process a newly fixated word. In addition, two articles mentioned the *immediacy assumption*, also established by Just and Carpenter [43,60], according to which a reader attempts to interpret each content word of a text immediately and without delay, even at the expense of assumptions that occasionally turn out to be wrong [97]. Of the articles examined, only Rouinfar *et al.* referred to another theory [1]—Henderson’s rubber band model of eye movements and attention [98,99].

F. Limitations

Several studies reported limitations when analyzing and interpreting eye-movement data. The following subsections summarize the most common limitations that were reported in extant literature.

1. Calling for additional (verbal) data to interpret eye movements

Numerous studies collected additional verbal data to infer meaning from students’ eye movements or to reduce noise in the data. For example, Van Gog *et al.* combined eye tracking with concurrent verbal protocols to shed more light on students’ troubleshooting performance [44]. Susac *et al.* used additional written explanations of multiple-choice test items to assess the correctness of student responses [64]. Studies without additional verbal or written data sources lacked interpretations and explanations, such as why information was not retained despite reading the text [62] or why students failed to make important transitions despite instructional guidance [43]. Moreover, several publications indicated that additional verbal data would be required to fully understand the results obtained via eye tracking [38,39,42,43,55,60,62,68].

One fear is that parallel speaking might influence the eye movements. The study of Ibrahim and Ding provides a methodical example to overcome this problem by processing the tasks once with and once without thinking aloud [48]. Thus, on the one hand, one measures the undistorted attention without the additional burden of thinking aloud in the first run and, on the other hand, still obtains the verbal data in the second run. However, this creates the problem that learners see the same stimulus twice in a short period of time. In this context, Klein *et al.* also indicated that eye-tracking data can be biased when the same or similar stimuli are viewed repeatedly [51]. Han *et al.* also expressed the concern that memorization and familiarity effects imply that pre- and post-testing is not easy; hence, they used different samples for pre- and post-test instead of matched samples to avoid students seeing the questions twice [45].

2. Low number of students

A desire to increase the sample size was expressed in many studies, which falls short of the size of questionnaire

studies due to the time-consuming nature of the data collection process [38–40,44,58,60,63,66]. Small sample sizes lead to lower statistical power; therefore, a few studies refrained from performing statistical significance tests and instead relied on the calculation of effect sizes [58] or on nonparametric procedures [43]. In assessment scenarios, the number of individuals who chose certain distractors shrinks further with small initial samples, thereby making group comparisons difficult [66]. In addition, small samples limit the ability to conduct mediator and moderator analyses, such as how reading comprehension affects the processing of text-picture combinations [60].

3. Stimuli restrictions

A few studies also reported limitations in terms of stimuli. For example, the studies presented only a limited range of stimuli [38,58], thereby causing the interpretations of eye movements to not be applicable to other types of graphs or charts. Low ecological validity was also mentioned with respect to the clinical nature of eye-tracking studies—for example, only a single example text was presented that students were required to understand [60].

It was also mentioned that certain stimuli must be analyzed individually, which leads to higher effort and lower generalizability [38,47]. This is the case, for example, when comparing different types of training that include moving and static images, which makes it difficult to define corresponding AOIs [47]. Thus, inconsistent item formats or different numbers of alternatives must be avoided when similar analysis procedures are required to be applied for every item [53,66]. On the other hand, individual analyses, for example of gaze paths, enable the provision of individualized feedback [46] and specificity would be lost if eye-tracking data were averaged over too many different stimuli [47].

Similarly, specificity would be lost if AOIs were selected too coarsely [42,62]. This can be remedied by qualitative analyses of heat maps [49] or exploratory analyses of AOI patterns [52,57], where attention can be viewed on small stimulus segments. Conversely, the studies of Madsen *et al.* and Rouinfar *et al.* revealed how relevant and irrelevant areas in the stimuli can be determined by preliminary studies through interviews [1,58,59]. This preliminary work made it possible to define high-resolution look zones that could be used to pinpoint differences in expertise.

4. Mobile eye tracking and real learning scenarios

Only one study uses eye-tracking glasses to capture visual attention in hands-on experiments [42]. While the high potential of this approach for typical learning scenarios in physics is apparent, Chien *et al.* reported that tremendous manpower is required for analyzing these data. The learners changed their visual space and moved objects from moment to moment; thus, the look zones must be adjusted manually frame by frame to derive accurate

results. This resulted in only a part of the data being analyzed. If automated gaze mapping in real-world scenarios would become possible through software solutions in the future, additional scenarios for physics learning could open up. This overcomes, for example, the limitation mentioned by Susac *et al.* whereby paper-and-pencil data cannot be captured with screen-based gaze tracking [63].

VIII. DISCUSSION

A. Dissemination of eye tracking in PER (RQ1)

1. Increasing use of eye tracking in PER

In this literature review, the current status quo regarding eye tracking in PER was examined and summarized. A thorough paper selection process yielded 33 studies that used eye tracking in physics education to investigate learning processes that were published in journals since 2005 and an additional 23 studies that were published as proceedings or book chapters. This number is significantly lower compared to mathematics education research [8], which shows the existing potential for eye tracking in PER. Furthermore, similar to previous reviews, there has been a substantial increase in the number of publications, particularly within the most recent years [6–8]. This confirms the ongoing popularity and importance of eye tracking; as explained by Strohmaier *et al.*, technological progress has made the implementation of studies easier and this leads to a greater awareness of fields beyond education [8]. The dynamic development of eye-tracking studies within physics education can also be gauged by comparing it to the development of eye-tracking methodology in general. For PER studies within the period 2014–2017, a relative increase of approximately 30% per year and within the period 2017–2020 of approximately 50% per year became evident, whereas it was 10% per year for the mentioning of eye tracking in general [100,101]. Although this result does not provide a qualitative argument for the validity of eye tracking in PER, it can be interpreted quantitatively as an argument in favor of the use of eye trackers in PER studies, as researchers have now recognized its potential and benefits.

2. Convenient use of eye tracking in assessment scenarios

Most of the research conducted thus far has concentrated on assessment scenarios. The importance of assessment in PER is reflected in the growing number of research-based distractor-driven multiple-choice items since the development of the FCI in the 1990s [30,102]; the further development of assessment is also considered as an important future research direction in PER [103]. Eye tracking enables access to process-related components in such scenarios without interfering with the process of information extraction and elaboration. Thus, eye tracking can provide the researcher with greater insight into the participants' problem-solving behavior, thinking, and approach

to solving physical problems. For example, a participant who answers correctly may still spend a significant amount of time analyzing incorrect choices, thereby suggesting that students have difficulty overcoming incorrect concepts [45]. Considering that concept inventories are continuously used for research purposes and researchers still put efforts into developing new instruments and modifying existing ones, visual attention analysis could reveal further information regarding the functionality of distractors, discrimination, and students' cognitive processes in problem solving at the test and single-item levels, which would overall promote instrument validation. The first steps in this direction of using eye tracking in terms of item and test analyses have already been mentioned and executed [45,48,53,55]. For example, Klein *et al.* confirmed item objectives by clustering eye-movement transitions in the context of kinematics graphs. Ibrahim and Ding stated that eye movements can reflect task demands on a procedural level well beyond classical measures in the context of synthesis problems [48]. The large number of studies in assessment scenarios can also be explained by the fact that these studies are rather suitable for eye-tracking data collection and comparatively easy to set up. The items can be presented as a whole on one screen, no scrolling is necessary, and no collaboration with other learners is required. Since eye trackers are designed to track the eye movements of a single participant sitting in front of a screen, they can be used to represent such scenarios very well. Hence, the good research situation in this scenario suggests that eye tracking will likely be increasingly used in assessment scenarios in the future.

3. Potential of eye tracking in learning scenarios and laboratory work

In contrast, the scenarios investigating knowledge construction by instructional designs or laboratory work were barely focused on. Here, the main focus was on learning physics using technological tools such as animations [47], simulations [42,43], visualization tools [1,59], or text-image combinations [40,54,60]. There is a broad base of eye-tracking studies in similar scenarios outside PER, summarized in several other reviews and not included here (e.g., Ref. [6]). In general, eye tracking offers the opportunity to contribute to an evaluation of how a particular learning material or method relates to higher or lower levels of learning. Further, specific added value for PER is evident, for example, in the research-based use of visual cues in conceptual tasks [1], physics-type work with measurement data [47], or the use of the predict-observe-explain method in demonstration experiments [57]. Eye tracking can be used to better understand how learners interact with simulations [43], how they use instructions [54], or how they read display elements in a simulation [42]. Eye-tracking data can help interpret the effectiveness or ineffectiveness of training, for example, by

revealing why students fail [43]. In addition, eye-tracking studies can be used to derive important practical implications for the design of textbooks and materials, such as which cues are displayed [1,54] or which visual stimuli are more comprehensible [40]. In perspective, eye tracking can be used as a diagnostic tool for adaptive learning systems (e.g., Ref. [104]) or serve as a learning method involved in targeted training of eye movements [105]. Only one study was identified with regard to laboratory work [42]. Process-based analysis of experimentation can reveal facets that are hidden from product-based analysis and can provide much insight into difficulties in experimentation, thus indicating the specific value of eye tracking in PER. Eye tracking can contribute to the clarification of physics education questions, for example, by studying the visual complexity of experimental setups [106], collaborative work and joint attention on relevant tasks [107], how this can be positively influenced, and which phases of experimentation cause difficulties. The lack of studies on laboratory work is probably due to the difficulty of setting up the eye-tracking environment for such tasks [42].

4. Focus on (multiple) representations with a broad coverage of physical topics

With regard to the studies' physical topics, a clear focus on learning of mechanics concepts became apparent—that is, kinematics, statics, forces, energy, and work. Mechanics is an essential basis for understanding the entire domain of the pursuing physics curriculum as it provides conceptual and mathematical foundations for all areas of physics [108]. However, the basis of this understanding is set in school and, thus, more studies using appropriate samples are needed (see Sec. VIII B). In addition to the aforementioned potential for eye tracking in general, perspectives for various areas of modern physics might be particularly promising, particularly those using one or more central representations—for example, quantum mechanics [109–111]. With respect to the educational topics, according to Ref. [30], eye-tracking research in PER focuses on examining the handling and choice of external representations (graphs, diagrams, motion maps, interference patterns, electrical circuits, vector fields, etc.). This emphasizes the scientific consensus that a gaze-data analysis “is very suited to study differences in attentional processes evoked by different types of multimedia and multirepresentational learning materials” [112] (p. 95). Since external representations are of great importance in physics education, eye tracking allows an analysis of the “behaviors that are difficult to articulate” [113] (p. 200) when dealing with representations and learning materials. Since eye movements reflect more than just consciously focused operations, gaze-data analysis reveals more complete information regarding the underlying mental processes [113] and, thus, can make an important contribution to the evaluation of material designs for teachers.

Furthermore, for multiple representations, there is sizable meta-analytic evidence in favor of a strong impact on learning (Cohen's $d = 0.75$) [114]. Based on these results, eye tracking, when it is used to inform teaching practice (beyond research purposes), can act with considerable educational “leverage”—that is, on instructional variables with established practical significance.

B. Methodological implementation of eye tracking in PER (RQ2)

1. High methodological standard of eye-tracking studies in PER

Going beyond the dissemination of the method, the second research question focused on the publications' specific methodological characteristics. In a meta-analysis on educational innovations in PER, Ruiz-Primo *et al.* identified the lack of attention given to the psychometric properties of the instruments used to measure learning outcomes as a major methodological threat to the validity of the studies [115]. This was not observed in the papers included in this study. With regard to the visual stimuli, it was found that many studies used pre-existing test materials to study attention, such as items from concept tests (e.g., TUG-K [49,55], FCI [45,56], R-FCI [46,50], CATs [37]) or items that had already been used in a similar manner in other studies (e.g., sets of problem pairs [38,63,64]). This reflects the long tradition and concurrent specificity in PER of developing and applying concept tests to analyze misconceptions and learning difficulties (see also Ref. [102]) and results in a high degree of standardization in empirical investigations in PER. Most studies quantitatively analyzed the eye movements on the visual stimuli by defining look zones, so-called AOIs. Only two studies used qualitative analysis approaches. Quantitative research methodology, in particular, satisfies the modern approach for improving education by using components of experimental scientific research practices based on objective data, disseminating results, and utilizing modern technology [116]. The predominance of this approach in the context of eye tracking shows the contemporary added value of the method in educational research and particularly in PER. Globally defined AOIs allowed the researchers to examine attention to large sections of the stimuli as well as gaze switches between them—for example, between question and options in multiple-choice tests. This was often used to examine stimuli characteristics and differences between student groups or to find correlations between performances, as discussed in the next section. On the other hand, locally defined AOIs enabled statements regarding the processing of the stimuli themselves, for example, to gain insights into relevant and irrelevant areas [58]. The results at the local level could be strongly linked to physics education implications—for example, the identification of misconceptions [44,53] or the targeted use of cues [1,59].

2. The challenge of small sample sizes

Considering the participants in eye-tracking studies, it was found that the studies in PER investigated larger samples than, for example, in mathematics education research ($M = 28.56 \pm 21.70$ participants [8]). Large samples imply a great technological and organizational effort, but they are highly desirable to aggregate statistically robust results, thereby enabling the highest possible statistical generalizability. A few studies highlighted sample size as an issue requiring improvements. To achieve larger samples, multiple eye trackers can be used in parallel to speed up data collection [40]. Indeed, Klein *et al.* as well as Susac *et al.* reported using four and three eye trackers in parallel, respectively [53,66]. However, despite the considerable evolution of eye trackers over the years, they are still not available in large quantities to a large number of researchers due to the high cost. Susac *et al.* addressed this issue and reported using inexpensive eye trackers with sufficient accuracy for the purpose of the study [66]. Another approach involves recruiting a large sample and performing eye tracking with only a subsample. Susac *et al.* showed that the test situation (eye tracking vs paper and pencil) did not influence students' test scores [63].

3. Prevalence of undergraduate students

With regard to the age group, the review revealed a dominance of introductory university students due to their availability, whereas approximately one-third of the studies were conducted with school students. Age is an important factor influencing eye movements [3], which is why attention must be paid to the fit of the learning material to the age of the target group. In this context, Strohmaier *et al.* expressed concerns regarding the transferability and generalizability of the results from students to pupils [8]. Considering the potential of eye tracking to serve school teachers in evaluating instructional materials, there is a need to focus more on this target group in future studies, in addition to the desire for larger samples.

4. Complete documentation of technological features is often lacking

As this review focuses on eye-tracking methodology, we also addressed the technological implementation of eye tracking. Creating a comfortable working environment for the subjects is important so that they can work with the study materials without being adversely affected. A balance must be achieved between the movement restrictions due to head or chin rests and the quality of the data. Note that when deciding on an eye-tracking device, price also plays a role when aiming for high precision and high temporal resolution in head-free conditions. Approximately half of the studies reported head-free conditions; in this case, a specification of measurement accuracy, for example, by

indicating the amount of captured eye positions, would be desirable to assess the accuracy of the data. It must be noted that possible exclusion criteria as well as eye-tracking devices, sampling rate, accuracy, and movement restrictions were not adequately reported in all studies. At this point, the importance of full documentation and reporting as stressed by Strohmaier *et al.* must be emphasized again [8]. As the authors state, this includes “a precise description of the apparatus including sampling rate and average accuracy; the existence or nonexistence of movement restrictions and information regarding the setup; the size of the stimuli; the distance between the stimuli and the participant; the monitors refresh rate; the calibration procedure and calibration accuracy threshold; the event detection algorithm and event detection thresholds; the position and size of any AOIs; the correlation between all used measures; and the amount of and reasons for data loss” [8] (p. 165).

5. The importance of additional data sources in interpreting gaze data

With regard to additional data, approximately two-thirds of the studies collected, for example, oral or written explanations from the subjects in order to triangulate them with the gaze data. This is a development that was already demanded by Lai *et al.*, as they called for establishing a link “between the micro-measurements (i.e., eye-tracking measurements) and the macro-measurements (i.e., performance or behavioral measurements)” [6] (p. 99) as a goal for future studies. Such a methodological triangulation enables approaching cognitive processing from different angles and, thus, helps to interpret eye-movement research in meaningful ways [105,117] as well as to improve confidence in results [118]. In multimedia learning, Hyönä calls for complementing eye-tracking data by off-line measures that index the outcome of learning in order to yield new insights into the multimedia processes [119]. Following eye-tracking research from other educational fields, methodological triangulation must particularly be used if the research question aims “to answer why people look where they look” [117] (p. 121) which applies, for example, if students' processing of representations is explored (e.g., Refs. [52,64]).

C. Findings, interpretation, and limitations of eye tracking in PER (RQ3)

Working through the identified studies, we inductively created four categories that structured the report on the findings. For a better orientation, we refer to Fig. 5, in which these categories are presented again with their characteristics. The following three sections (VIII C 1–VIII C 3) refer to the upper left square—that is, the performance-based analysis of eye-tracking data. Thereafter, the remaining results are summarized.

	Performance-based <ul style="list-style-type: none"> ▪ Linking visual attention to response accuracy, overall performance, or expertise ▪ Exploring perceptual mechanisms underlying expert performance during the process of problem solving ▪ The definition of AOIs and selection of eye-tracking measures is important for performance discrimination. 	Stimuli-based <ul style="list-style-type: none"> ▪ Comparing visual attention for a variation of the visual stimulus ▪ Visual processing of different task requirements, for example, when adding or highlighting information ▪ Provide guidance for the (re)design of new or existing test tasks or learning materials 	
	Progression-based <ul style="list-style-type: none"> ▪ Exploring changes of visual attention over time ▪ Studying the impact of interventions on procedural knowledge (information extraction, focusing relevant areas) ▪ Exploring students' interaction with learning materials in greater detail 	Student-based <ul style="list-style-type: none"> ▪ Comparing different groups of students to study how different groups of learners deal with learning materials ▪ Identification of discipline-specific characteristics and ways of thinking ▪ Enriching our understanding of visual expertise using <i>a priori</i> defined contrasts 	

FIG. 5. Overview of the most frequent independent variables for which the eye-tracking metrics have been evaluated and their characteristics.

1. *The connection between accurate responses and time-aggregated eye-movement measures becomes apparent when the visual stimulus is partitioned meaningfully*

Our analysis revealed a clear focus of PER eye-tracking studies on a performance-based comparison of gaze data between correct and incorrect answers and between experts and novices. Previous reviews in educational research found a lack of evidence for the relationship between eye-movement measures and student (learning) performance [6]. However, our findings suggest an anticipation in the field of PER in recent years for eye tracking to be centrally used in assessment scenarios; accordingly, research on competency differences is emphasized, which is rather similar to other educational settings, for example, medical education [120,121]. Researchers see the added value of using eye-tracking methodology for this purpose in gaining “significant insights into some of the perceptual mechanisms underlying expert performance” [5] (p. 524). However, a closer examination of the study results reveals that the choice of eye-movement measurements and the definition of AOIs is crucial in this line of research [119]. A consistent result of the studies reviewed here was the lack of relationship between response accuracy and total viewing time or number of fixations during an entire task. To elucidate performance differences using eye-tracking metrics, a finer analysis of attentional distribution is required, which could be achieved, for example, by finer segmentation of the stimulus or the use of progression-based measures. Thus, in order to draw valid inferences from eye-tracking data, a sensitive, theory-driven experimental design and operationalization of the dependent variables—that are, eye-tracking measures and AOIs, is required. Further, when viewing times or fixation counts were compared for relevant and irrelevant areas, a correlation with performance was consistently found in all the studies. Returning to the previous research question, this segmentation of stimuli into smaller sections was based either on

surface features (global AOIs) or on educational considerations (e.g., specific areas in a diagram; local AOIs), which provided insights into representation-specific misconceptions and processing strategies (local AOIs).

Further, by creating a stronger contrast by comparing high-performing and low-performing students (experts and novices), the above-mentioned results become even clearer. Here, there are consistent results across contexts that reveal that high-performing students focus their attention more quickly on relevant parts of the stimuli, spend more time on relevant areas, and have longer average fixation durations during problem solving than low-performing students. This finding is well in line with the results reported by Gegenfurtner *et al.*, thereby confirming the information reduction hypothesis, which proposes that “expertise optimizes the amount of processed information by neglecting task-irrelevant information and actively focusing on task-relevant information” [5] (p. 539f).

At this point, the sensitivity of the eye-tracking method became especially clear as differences in the evaluation and interpretation of the gaze data (e.g., AOI definition) produced seemingly contradictory results. Similar effects were reported by Gegenfurtner *et al.* as well as Strohmeier *et al.* as they found optimized information processing to be associated, on the one hand, with faster information extraction (and, thus, less time and fixations on important areas) and, on the other hand, with relatively more time on relevant than on irrelevant areas [5,8]. Thus, this emphasized the need for a sensitive, theory-driven experimental design as well as a well-considered, critical evaluation once again.

2. *Transition measures have been most informative in learning scenarios*

Some of the studies considered here used other measures in addition to the number of fixations and dwell times, which is generally strongly advocated in educational

research because these measures allow a more detailed picture of cognitive-visual behavior [119]. For example, gaze transitions have been analyzed with respect to performance. Here, no clear correlation emerged for assessment scenarios, but for knowledge construction processes. For the latter, the reported results fit the current body of research—that is, that a higher number of transitions between different elements in multimedia learning environments is associated with better learning performance [122–124]. This is also supported by the learning theory, since according to the cognitive theory of multimedia learning by R. E. Mayer [125] transitions among elements reflect mental integration processes [7]. On the other hand, in assessment scenarios, erratic behavior between questions and options or between different options may indicate uncertainty, although integration mechanisms between different elements of a task, such as stem and graph, might be necessary. More systematic studies are required in this regard to understand the role of gaze transitions in assessment scenarios. There is a limited number of studies in PER that go beyond these metrics and investigated mean fixation duration and saccade length in the context of performance.

3. Time-based measures provide information regarding response confidence

In addition to response accuracy, certain studies also used students' confidence as a split criterion for their analyses. In educational research, confidence judgments are well recognized as being an important part of knowledge assessment [126] since they include the subjects' metacognitive abilities [127]. As our analysis has revealed, there is consensus that confident answers were correlated with shorter viewing times. Thus, the combination of eye tracking, response accuracy, and response confidence can be a fruitful method to test knowledge more thoroughly [126] and to elucidate the connection between metacognitive abilities and content knowledge.

4. The comparison of visual attention across different stimuli informs material development

More than half of the studies investigated gaze behavior between different stimuli, for example, by adding a form of representation, contrasting two different forms of representation or task, or teaching different strategies. On the one hand, this line of research succeeds in identifying specific properties of the stimuli that would remain hidden without eye tracking—for example, the visual processing of different diagram concepts. On the other hand, it also allows tracing back instructional effects to the visual processing of the stimuli. This supports the statement by Van Gog and Scheiter that “for research on multimedia multi-representational learning materials eye tracking can provide unique information concerning what medium or representations are visually attended to, in what order, and for how long” [112] (p. 59). As a result, the analysis of eye

movements enables a detailed understanding of the material and the triggered cognitive processes, indicates misconceptions, and can provide guidance for the (re)design of new or existing test tasks or learning materials. However, analyzing visual attention patterns of correct problem solvers in certain FCI items revealed that students' visual strategies are strongly representation dependent [56]. Thus, viewing patterns associated with correct responses cannot be generalized for questions with different representations. In fact, gaze behavior for graphs is strongly dependent on the context in which it is presented.

5. The development of attention over time enables conclusions to be drawn regarding learning effects

The progression-based analysis of eye movements sheds light on the impact of teaching interventions on changes in students' visual attention. As demonstrated in the studies, this information can be used to explore the result of learning—that is, strengthening expertise—on a process level in addition to test scores. This is particularly useful when instructions aim at representational competencies that require information extraction from visual stimuli. Here, changes of how learners interact with the stimuli can be explored. This might also be helpful when interventions fail. In this context, visual information can provide information on whether the problem-solving behavior was influenced (possibly for the worse) by an instruction.

Apart from comparing visual attention between two or more time points, the continuous progression of visual attention over time when students engage with learning material was also analyzed. In addition to a global static analysis (which might not be able to capture the learning process in minute detail [119]), a fine-grained analysis of eye-movement measures over time can reveal the point at which students fail.

6. Comparing visual attention among different student groups

Studies that compare gaze data among different groups of students aim to identify discipline-specific characteristics and ways of thinking, which can, in turn, inform material development. The advantage of examining research questions that establish a contrast based on external criteria that are fixed *a priori* (such as expertise, spatial abilities, etc.) is to obtain conclusions that go beyond correlative relationships.

For example, spatial abilities are attributed a crucial role in physical problem solving and reasoning [128,129] and eye tracking can reveal the differences in visual behavior between students with high and low spatial ability. In this manner, we obtain information on how different groups of learners deal with learning material in order to subsequently generate intervention measures that are appropriate for the target group.

By comparing experts and novices or physicists and nonphysicists, important insights are gained regarding

the different levels of information extraction, information processing, and problem solving, which, in turn, enrich our understanding of visual expertise and provide directions for improvement to those who need them.

7. Pragmatic justifications for the use of eye tracking predominate in PER eye-tracking studies

Most of the research examined in this paper adopted a pragmatic viewpoint and assumed that there is a link between visual attention and cognitive mechanisms or, alternatively, cited prior work that has found such a link. While this approach is not under criticism, it does impede a theory-driven selection of eye-tracking metrics that goes beyond the distribution of attention as aggregated time or count measures. In their meta-analysis, Alemdag and Cagiltay reveal which cognitive processes in knowledge acquisition can be operationalized by which eye-tracking measures [7]. Now, this very recent work was not available to most researchers, and it is also unclear whether these operationalizations are valid for assessment scenarios. As eye movements reflect and are influenced by cognitive processes, but cognitive processes cannot be directly inferred from eye-tracking data [105], we wish to emphasize the general awareness of considering theoretical frameworks when setting up eye-tracking studies. Following Jarodzka *et al.*, for an appropriate interpretation

of eye-tracking data, theoretical approaches and models always need to be the basis for designing experiments as well as for analyzing and interpreting eye-tracking data [105].

8. Addressing the limitations of eye-tracking studies

This synthesis clearly demonstrated the added value that can be expected when using eye-tracking methodology in PER. However, the analysis of eye movements also has clear limitations and restrictions with regard to its application in research and the interpretation of the data, as listed in Sec. VII F. We summarize these limitations and possible ways to handle them in Table VI.

Therefore, for an appropriate and thoughtful use of the method as well as the development of an understanding of what can be derived from eye-movement data, an awareness and understanding of these limitations is crucial [117]. In particular, before conducting one's first eye-tracking study, the limitations must be reflected upon and incorporated into the experimental design and research questions. At this point, however, it must be added that there were only few qualitative studies that evaluate and compare the above-mentioned approaches with regard to coping with particular limitations. Regarding the collection of additional verbal data, emerging studies compare the use of two different methods—that are, concurrent and retrospective think aloud with eye tracking [130]. In an early

TABLE VI. Methodological limitations of eye tracking and how researchers have dealt with them.

Limitation	Actions or measures	Conflicts or challenges
Ambiguity in the interpretation of eye-movement data → collection of additional qualitative data	Concurrent think aloud (verbal) data Present stimuli twice, once with and without verbal data Retrospective think aloud Combining multiple-choice answers with written explanations	Cognitive burden—that is, thinking and speaking (eye movements might be influenced by the cognitive processes regarding verbal articulation) Memorization and familiarity effects during the second run (use different subsamples) Requires training and adequate questions of the investigator to describe eye movements Reduces noise in the performance data but does not inform interpretation of eye movements
Small number of students	Use eye trackers in parallel to speed up data collection Use low-cost eye trackers Using eye tracking only for a subset of the sample	High cost of accurate eye-tracking systems, requires more researchers, challenge of merging the datasets Accuracy issues Incomplete datasets, different test situations
Ecological validity suffers from stimuli restrictions and setting	Keep the setting as natural as possible (head-free tracking) Pilot studies for optimizing visual stimuli	Accuracy issues More time required
No access to visual attention during real-world scenarios (lab work, paper and pencil) with screen-based eye tracking	Using mobile eye-tracking glasses	Tremendous manpower required for data analysis (option for short measurements)

methodological study in PER, Van Gog *et al.* studied the amount of problem-solving process information that could be uncovered using concurrent and retrospective think aloud without gaze data and cued retrospective think aloud with eye tracking [131]. Thus, there is a need for further research in this area on a methodological level, especially for PER-specific questions. The subsequent presentation of stimuli in combination with eye tracking, as conducted by Ibrahim and Ding, represents another alternative in this research direction that must be evaluated in comparison with other methods. Regarding the other limitations, such as the use of eye tracking with only a subset of the sample, there are also no systematic studies in PER.

IX. FUTURE DIRECTIONS

As our work has revealed, eye-tracking technology is capable of revealing learning scenarios that describe the underlying psychological processes of different learning goals in physics education research—that is, investigating assessment, knowledge acquisition, and laboratory work. However, there is an imbalance in the number of studies that have been conducted in these areas and this suggests that more efforts are needed to explore, for example, laboratory work. Experimentation plays a central role in physics education, particularly at the action level, and enables collaboration among learners. Here, eye tracking can help to obtain clues to student difficulties in experimentation through visual attention. The visual perception of (demonstrated) experiments with different setups can also be helpful both for physics in the lecture hall and for teaching physics online. Whether and to what extent collaborative experimentation is successful compared to isolated learning can be investigated by process-based analysis, like eye tracking with glasses. Joint visual attention plays a significant role in this, as Schneider and Pea found that awareness of visual attention can enhance collaborative colleagues' joint visual attention and improve learning gains [132]. Since the effort for eye-tracking studies with glasses is very high and time consuming, methodological steps must be taken to investigate meaningful experimental settings in a controlled manner.

As the research has shown, several cognitive processes (e.g., expert and novice thinking, learning difficulties) can be identified by analyzing visual attention. This information can be used to develop adaptive learning systems that take the learners' level of knowledge into account. Adaptive learning systems equipped with eye-tracking technology can provide dynamic learning aids for all students to enable better learning. For example, a system can respond adaptively when the viewer looks at certain areas in a diagram that are off-target or related to a misconception. Although this may appear technically possible, the extent to which such adaptive systems actually improve the learning process needs to be tested empirically.

In general, visual-spatial skills play an important role in physical problem solving and reasoning, particularly when

(multiple) representations have to be interpreted. This ability has already been considered in a few studies, but it appears useful to investigate the relationship between eye movements and spatial abilities in more detail in the future. There is potential for deeper insights into the facets of cognitive-spatial skills and their link to cognitive-visual processes. Here, targeted training of eye movements (e.g., through modeled gaze paths and eye-movement modeling examples [105]) can help low-spatial ability students to target their gaze to relevant areas or to force the reading of a representation. This dual role of eye tracking—as a research and learning tool—is also evident in the presentation of the students' own view paths for reflecting on their learning. Therefore, eye tracking can make conscious and unconscious processes of learners visible when dealing with digital learning materials—for example, how they perceive simulations or animations. This allows conclusions to be drawn regarding design principles and a criterion-guided selection of learning materials.

Another trend emerging from recent work is the use of eye tracking for task analysis and validation in order to enrich product-based analysis techniques (analysis of test scores) with behavioral measures through visual attention. In particular, the expansion to more complicated tasks compared to single-concept problems (e.g., synthesis problems, multiple representations) make it worthwhile to learn more about this underresearched type of problems.

X. CONCLUSION

As our study results reveal, an increasing number of physics education researchers are using eye tracking to learn more about physics learning by analyzing students' visual attention. Eye tracking was most commonly used in representation-rich assessment scenarios to establish a link between visual attention and expertise. Our analysis found that this is most successful when extreme groups (high vs low achievers) are considered or when the analysis considers the spatial separation of relevant and irrelevant information (local AOIs). In this manner, known misconceptions (e.g., in the interpretation of diagrams) were further validated or unknown ones were observed. Eye tracking has also enriched physics education research by providing additional information on the test tasks themselves, the cognitive processes they evoke, and how different tasks differ on a process level. Moreover, in learning scenarios, the effectiveness of visual cues or different multimedia presentations for physics understanding could be better investigated and interpreted through eye tracking.

Eye tracking has the potential to provide new insights into physics problem solving and learning. Given the growing number of studies in which eye-movement measurement is already being applied, we hope that this review can guide future researchers in this field and assist them in using eye tracking in a reflective manner.

ACKNOWLEDGMENTS

This material is based on work supported by the Joachim Herz Foundation under Project No. 850030 (“eyePhysics”). We further acknowledge support by the Open Access Publication Funds of the Göttingen University. We are very grateful for the thorough comments from Prof. Andreas Müller (Université de Genève) in the preparation of our story and from three anonymous reviewers during the review process.

APPENDIX

Tables VII–X contain the information about the eye-tracking studies under consideration in this systematic

review. Specifically, Table VII contains the information about the dissemination of the eye-tracking studies referring to RQ1 (i.e., publication year, scenario, topical area, and physics domain). Note that both journal publications and secondary literature are listed there.

Further, Tables VIII and IX belong to RQ2 and contain information about the methodological implementation of eye tracking (i.e., stimuli used in the studies, areas of interest segmentation, sample, sample size, device, eye-tracking calibration information, accuracy, and additional data).

Finally, Table X addresses the aspects of the third research question (i.e., eye-tracking measures, theoretical background, and data split).

TABLE VII. Dissemination of eye-tracking studies in physics education research: RQ1.

Publication	Scenario ^a	Topical area ^b	Physics domain
<i>Journal articles</i>			
Bi and Reid [37]	Assessment (S)	Conceptual understanding	Statics
Brückner <i>et al.</i> [38]	Assessment (S)	Graph understanding	Kinematics
Brückner <i>et al.</i> [39]	Assessment (L)	Graph understanding	Kinematics
Chen <i>et al.</i> [40]	Knowledge construction	Learning from text w. Rs.	Electricity (electric circuits, electric potential difference)
Chen <i>et al.</i> [41]	Assessment (S)	Conceptual understanding	Various (mechanics, optics, electromagnetics, and thermodynamics)
Chien <i>et al.</i> [42]	Laboratory work	Experimentation	Thermodynamics (ideal gas law)
Chiou <i>et al.</i> [43]	Knowledge construction	Computer-based learning	Optics (refraction of light)
Van Gog <i>et al.</i> [44]	Assessment (S)	Representations	Electrical circuits
Han <i>et al.</i> [45]	Assessment (L)	Conceptual understanding	Forces
Hejnová and Kekule [46]	Assessment (S)	Representations	Forces
Hoyer and Girwidz [47]	Assessment (L)	Representations	Coordinate systems
	Knowledge construction	Computer-based learning	
Ibrahim and Ding [48]	Assessment (S)	Conceptual understanding	Various (Synthesis problems, mechanics)
Kekule [49]	Assessment (S)	Graph understanding	Kinematics
Kekule and Viiri [50]	Assessment (S)	Representations	Mechanics (forces)
Klein <i>et al.</i> [51]	Assessment (L)	Representations	Vector fields
Klein <i>et al.</i> [52]	Assessment (S)	Graph understanding	Kinematics
Klein <i>et al.</i> [53]	Assessment (S)	Graph understanding	Kinematics
Klein <i>et al.</i> [54]	Assessment (L)	Representations	Vector fields
	Knowledge construction	Learning from text w. Rs.	
Klein <i>et al.</i> [55]	Assessment (S)	Graph understanding	Kinematics
Kozhevnikov <i>et al.</i> [56]	Assessment (S)	Conceptual understanding, graph understanding	Kinematics, forces
Küchemann <i>et al.</i> [57]	Assessment (L)	Representations	Coordinate systems, inertial forces
Madsen <i>et al.</i> [58]	Assessment (S)	Conceptual understanding, graph understanding	Energy, kinematics, motion
Madsen <i>et al.</i> [59]	Assessment (A)	Conceptual understanding, graph understanding	Energy, kinematics, motion
	Knowledge construction	Computer-based learning	
Mason <i>et al.</i> [60]	Knowledge construction	Learning from text w. Rs.	Forces (inclined plane)
Rouinfar <i>et al.</i> [1]	Assessment (A)	Conceptual understanding, graph understanding	Energy, kinematics, motion
	Knowledge construction	Computer-based learning	

(Table continued)

TABLE VII. (Continued)

Publication	Scenario ^a	Topical area ^b	Physics domain
Škrabánková <i>et al.</i> [61]	Assessment (S)	Graph understanding	Various (kinematics, oscillations and waves, molecular physics)
Smith <i>et al.</i> [62]	Knowledge construction	Learning from text w. Rs.	Mechanics
Susac <i>et al.</i> [63]	Assessment (S)	Representations	Measurement uncertainty
Susac <i>et al.</i> [64]	Assessment (S)	Graph understanding	Kinematics
Susac <i>et al.</i> [65]	Assessment (S)	Representations	Energy
Susac <i>et al.</i> [66]	Assessment (S)	Representations	Wave optics
Susac <i>et al.</i> [67]	Assessment (S)	Representations	Wave optics
Wu and Liu [68]	Assessment (S)	Representations	Kinematics
<i>Conference Proceedings and book chapters</i>			
Agra <i>et al.</i> [69]	Assessment (A)	Conceptual understanding, graph understanding	Energy conservation, speed
	Knowledge construction	Computer-based learning	
Holubová <i>et al.</i> [70]	Assessment (S)	Conceptual understanding	Forces
Kekule [71]	Assessment (S)	Graph understanding	Kinematics
Klein and Kuhn [72]	Assessment (L)	Representations	Vector fields
Klein <i>et al.</i> [73]	Assessment (S)	Representations	Electrostatics
Klein <i>et al.</i> [74]	Assessment (S)	Representations	Vector fields
	Knowledge construction	Learning from text w. Rs.	
Küchemann <i>et al.</i> [75]	Assessment (S)	Graph understanding	Kinematics
Madsen <i>et al.</i> [76]	Assessment (S)	Conceptual understanding	Various topics, mechanics (motion)
Madsen <i>et al.</i> [77]	Assessment (S)	Conceptual understanding, graph understanding	Various (mechanics (motion, energy conservation, torque, pendulum), electrostatics (parallel plate capacitors))
Mozaffari <i>et al.</i> [78]	Assessment (S)	Representations	Kinematics (trajectory, air resistance, motion)
Mozaffari <i>et al.</i> [79]	Assessment (L)	Representations	Vector fields
Mozaffari <i>et al.</i> [80]	Assessment (L)	Representations	Vector fields
Mozaffari <i>et al.</i> [81]	Assessment (S)	Representations	Kinematics (trajectory, air resistance, motion)
Ohno <i>et al.</i> [82]	Assessment (S)	Conceptual understanding	Forces
Rosengrant <i>et al.</i> [83]	Assessment (S)	Representations	Electrical circuits
Rosengrant [84]	Assessment (S)	Representations	Electrical circuits
Rosiek <i>et al.</i> [85]	Assessment (S)	Conceptual understanding	Forces
Rosiek <i>et al.</i> [86]	Assessment (S)	Graph understanding	Kinematics
Rouinfar <i>et al.</i> [87]	Assessment (A)	Conceptual understanding, graph understanding	Energy, kinematics, motion
	Knowledge construction	Computer-based learning	
Rouinfar <i>et al.</i> [88]	Assessment (A)	Conceptual understanding, graph understanding	Energy, kinematics, motion
	Knowledge construction	Computer-based learning	
Viiri <i>et al.</i> [89]	Assessment (S)	Representations	Forces
Viiri <i>et al.</i> [90]	Assessment (S)	Representations, conceptual understanding	Forces
Wu <i>et al.</i> [91]	Assessment (A)	Conceptual understanding, graph understanding	Energy, kinematics, motion
	Knowledge construction	Computer-based learning	

^aS = Single point (one time) assessment, L = longitudinal assessment, A = adaptive assessment.

^bLearning from text w. Rs. = Learning from text with representations.

TABLE VIII. Stimuli used in the studies, areas of interest, and sample: RQ2.

Publication	Stimuli information ^a	AOI segmentation ^b	Sample ^c
Bi and Reid [37]	CATS items	Global (question, diagram, options)	20 U
Brückner <i>et al.</i> [38]	Four isomorphic pairs of linear graph tasks	Global (question, graph, options), local (graph, axes, axis labels)	23 U
Brückner <i>et al.</i> [39]	Four isomorphic pairs of linear graph tasks	Global (task question, the graph itself, options), local (axes, graph, axis labels)	41 U
Chen <i>et al.</i> [40]	Online electricity learning program	Global (base and target domains), local (pump, water wheel, water-level difference)	80 S
Chen <i>et al.</i> [41]	Computer-based physics concept questions	Global (whole presentation), local (picture, text)	63 U
Chien <i>et al.</i> [42]	Experiment in simulation-based or microcomputer-based laboratory, worksheet	Global (worksheet, experimental zone)	50 S (7 ^d)
Chiou <i>et al.</i> [43]	Physics simulation interface	Global (up panel, question, down panel, light, reflection, and refraction)	51 S (24 ^d)
Van Gog <i>et al.</i> [44]	Troubleshooting tasks (malfunctioning circuits, offered in a simulation program)	Global (whole circuit), local (battery, lamp and switch)	26 M (11 ^d)
Han <i>et al.</i> [45]	Full test: FCI in a web-based interface	Global (narrative text, picture (if any), choices, and other areas)	89 U
Hejnová and Kekule [46]	R-FCI items	No AOIs defined (qualitative analysis)	34 M (11 ^e)
Hoyer and Girwidz [47]	Eight assignment task (transfer data to diagram); dynamic animation	Global (one AOI with experimental setting and diagram)	119 S (6 ^e)
Ibrahim and Ding [48]	Four synthesis problems (sequential and simultaneous problems)	Global (problem statement, diagram), local (depicting entities and labels in diagram)	22 U
Kekule [49]	TUG-K items	No AOIs defined (qualitative analysis)	25 M
Kekule and Viiri [50]	R-FCI items	Global (question, options)	46 M (11 ^e)
Klein <i>et al.</i> [51]	Eight vector field plots	Global (whole vector field), local (axes)	41 U
Klein <i>et al.</i> [52]	Four isomorphic pairs of linear graph tasks	Global (question, graph, options), local (axis and tick labels, AOI pattern)	69 U (2 ^e)
Klein <i>et al.</i> [53]	Full test: TUG-K	Global (questions, options)	115 S
Klein <i>et al.</i> [54]	Initial vector field, instruction pages, 8 vector field plots	Global (definition, strategy, vector field, hint)	32 U
Klein <i>et al.</i> [55]	Full test: TUG-K	Global (questions, options)	115 S
Kozhevnikov <i>et al.</i> [56]	Mechanics diagnostic test items and FCI items	Local (axes, axis-labels, line segments within the graph)	15 S
Küchemann <i>et al.</i> [57]	Multiple-choice tasks with trajectories as options	Local (AOI pattern covering the figure of the item stimulus)	21 U
Madsen <i>et al.</i> [58]	Ten multiple-choice conceptual physics problems (with diagram)	Global (whole diagram), local (relevant AOIs, salient AOIs, and novicelike AOIs)	24 M
Madsen <i>et al.</i> [59]	Four sets of related conceptual introductory physics problems (with diagram)	Local (AOI pattern, expertlike, and novicelike AOI)	63 U (4 ^{d,e})
Mason <i>et al.</i> [60]	Information text (with an abstract/concrete illustration)	Global (text, illustration), local (topic text sequence, medial text sequence)	74 S (14 ^e)
Rouinfar <i>et al.</i> [1]	Four sets of related conceptual introductory physics problems (with diagram)	Local (AOI pattern, expertlike, and novicelike AOI)	80 U
Škrabánková <i>et al.</i> [61]	Seven problem-solving tasks with graphs	Global (whole stimuli, task, options)	52 S (12 ^e)
Smith <i>et al.</i> [62]	Worked-out-example with text and math	Global (textual and mathematical information), local (smaller granularity)	43 U
Susac <i>et al.</i> [63]	Eight multiple-choice test items about measurement uncertainties (PMQ)	Global (introduction, data, question, graphical representation, all)	30 U
Susac <i>et al.</i> [64]	Four isomorphic pairs of linear graph tasks	Global (question, graph, options), local (axes labels, graph)	90 U
Susac <i>et al.</i> [65]	Six multiple-choice questions on energy	Global (question, options, diagram, all)	60 U
Susac <i>et al.</i> [66]	Four multiple-choice questions on interference and diffraction patterns	Global (question, options)	35 S
Susac <i>et al.</i> [67]	Seven multiple-choice questions on interference and diffraction patterns	Global (question, options, patterns)	52 S

(Table continued)

TABLE VIII. (Continued)

Publication	Stimuli information ^a	AOI segmentation ^b	Sample ^c
Wu and Liu [68]	Three argumentation evaluation tasks and one argumentation generation task	Global (question stem, table, figure, equation)	96 U

^aConcept Assessment Tool for Statics (CAS), Test of Understanding Graphs in Kinematics (TUG-K), Force Concept Inventory (FCI), Representational variant of the FCI (R-FCI), Physics measurement questionnaire (PMQ).

^b(cf. Sec. V).

^cUndergraduates (U), School (S), Mixed (M).

^dExcluded based on selection criteria.

^eExcluded due to technical issues.

TABLE IX. Device, eye-tracking procedure information, and additional data: RQ2.

Publication	Device ^a	Calibration ^b	Accuracy ^c	Additional data ^d
Bi and Reid [37]	Tobii X (60)	HF	R	S, V, other
Brückner <i>et al.</i> [38]	Tobii X3 (120)	HF, 60 cm, 9-point, I-VT		
Brückner <i>et al.</i> [39]	Tobii X3 (120)	9-point, I-VT	0.4 deg	C
Chen <i>et al.</i> [40]	EyeLink (1000)	9-point		
Chen <i>et al.</i> [41]	EyeLink (1000)	52 cm, 9-point		
Chien <i>et al.</i> [42]	FaceLAB, MobileEye (30)	HF, I-VT		
Chiou <i>et al.</i> [43]	Eye tribe (30)	CR, 68 cm		
Van Gog <i>et al.</i> [44]	SMI (50)	FHR, 70 cm, I-DT	0.5 deg	V, mental effort
Han <i>et al.</i> [45]	SMI RED (60)	HF, 50 cm, 9-point	0.5 deg	
Hejnová and Kekule [46]	Tobii TX (300)	70 cm, 5-point, I-VT	0.5 deg	V
Hoyer and Girwidz [47]	LC eye follower (60)	HF, I-DT	0.4 deg, TA	S
Ibrahim and Ding [48]	EyeLink (1000)			V
Kekule [49]	Tobii TX (300)	70 cm, 5-point, I-VT	0.5 deg, R	V, other
Kekule and Viiri [50]	Tobii TX (300)	5-point, I-VT	0.5 deg, R	V
Klein <i>et al.</i> [51]	Tobii X3 (120)	HF, 61 cm, 9-point, I-VT	0.4 deg, R	V
Klein <i>et al.</i> [52]	Tobii X3 (120)	HF, 9-point, I-VT	0.4 deg	C
Klein <i>et al.</i> [53]	Tobii X3 (120)	HF, 9-point, I-VT	0.4 deg	C
Klein <i>et al.</i> [54]	Tobii X3 (120)	I-VT		C, S, mental effort, other
Klein <i>et al.</i> [55]	Tobii X3 (120)	62 cm, 5-point, I-VT	0.4 deg	C
Kozhevnikov <i>et al.</i> [56]	SMI iView-X RED (n.a.)	CR and FHR, 9-point	R	S
Küchemann <i>et al.</i> [57]	Tobii X3 (120)		0.4 deg	C, V
Madsen <i>et al.</i> [58]	EyeLink (1000)	FHR, 61 cm, 9-point, I-VT	0.5 deg, TA	V
Madsen <i>et al.</i> [59]	EyeLink (1000)	FHR, 61 cm, 9-point, I-VT	0.5 deg	V
Mason <i>et al.</i> [60]	Tobii T (120)	I-VT		S, other
Rouinfar <i>et al.</i> [1]	EyeLink (1000)	CR and FHR, 61 cm, 9-point	0.5 deg, TA	V
Škrabánková <i>et al.</i> [61]	GazePoint 3 (60)	I-DT	R	
Smith <i>et al.</i> [62]	Tobii 1750 (n.a.)	HF		
Susac <i>et al.</i> [63]	SMI iView HiSpeed (500)	HF, 50 cm, 13-point, O	0.25–0.5 deg	V
Susac <i>et al.</i> [64]	SMI iView HiSpeed (500)	HF, 50 cm, 13-point, O	0.25–0.5 deg	Paper pencil data, other
Susac <i>et al.</i> [65]	SMI iView HiSpeed (500)	HF, 50 cm, 13-point, O	0.25–0.5 deg	
Susac <i>et al.</i> [66]	SMI RED (120)	HF, 50 cm, 5-point, I-DT	R	
Susac <i>et al.</i> [67]	SMI RED (120)	HF, 50 cm, 5-point, I-DT	R	Paper pencil data
Wu and Liu [68]	Tobii X3 (120)	FHR, 70 cm, 9-point		V

^aFrequency in Hz in parentheses.

^bInformation on movement restriction [head-free (HF), chin rest (CR), forehead rest (FHR)], distance between participant and screen, *n*-point calibration (where available), and fixation identification algorithm [velocity-threshold identification (I-VT), dispersion-threshold identification (I-DT), other (O)].

^cSpatial resolution of device in degree visual angle and information about accuracy [threshold agreement (TA), reported otherwise (R)].

^dConfidence (C), spatial ability (S), verbal data (V).

TABLE X. Eye-tracking measures and variables used in the studies: RQ3.

Publication	ET measure ^a	ET theory ^b	Data analysis (data split) ^c
Bi and Reid [37]	TFD, FC	EMA	Performance (correct vs incorrect)
Brückner <i>et al.</i> [38]	FC, TC	EMA	Performance (correct vs incorrect)
Brückner <i>et al.</i> [39]	TFD	EMA	Stimuli (task type), student (physics vs economics), progression (pre vs post)
Chen <i>et al.</i> [40]	TFD, other		Stimuli (pictorial vs text, analogies vs metaphors), performance (high vs low performers)
Chen <i>et al.</i> [41]	MFD, SL, other	EMA	Stimuli (picture vs text), Performance (correct vs incorrect), post-hoc progression (duration of n-th fixation)
Chien <i>et al.</i> [42]	FC, MFD	EMA	Stimuli (laboratory setting), progression (worksheet sections), performance (high vs low achievers)
Chiou <i>et al.</i> [43]	TFD, FC, MFD, other	IA, EMA	Performance (correct vs incorrect), post-hoc progression (different simulation phases)
Van Gog <i>et al.</i> [44]	TFD, FC, MFD, TC		Progression (problem phases), performance (high vs low expertise)
Han <i>et al.</i> [45]	TFD	EMA	Student = progression (pre vs post), performance (option choice, high vs low achievers)
Hejnová and Kekule [46]	none		Qualitative analysis (performance; high vs low achievers)
Hoyer and Girwidz [47]	FC, SA, SL	EMA	Stimuli (trainings condition)
Ibrahim and Ding [48]	TFD, TC	EMA	Stimuli (sequential vs simultaneous tasks), performance (correct vs incorrect)
Kekule [49]	None		Qualitative analysis; heat maps split based on performance (correct vs incorrect, high vs low achievers)
Kekule and Viiri [50]	TFD	EMA	Stimuli (test version), performance (correct vs incorrect, high vs low achievers)
Klein <i>et al.</i> [51]	FC, MFD, SA, SL	EMA	Stimuli (strategy), performance (correct vs incorrect, high vs low achievers)
Klein <i>et al.</i> [52]	TFD, FC, SA		Stimuli (task type), student (physics vs economics)
Klein <i>et al.</i> [53]	TFD		Performance (high vs intermediate vs low confidence, correct vs incorrect, achievement level)
Klein <i>et al.</i> [54]	TFD, FC, TC, SA		Stimuli (cues vs no cues), progression (pre vs post), performance (high vs low achievers, correct vs incorrect), post-hoc progression (time intervals)
Klein <i>et al.</i> [55]	TC		Stimuli (task type), performance (correct vs incorrect, high vs low confidence)
Kozhevnikov <i>et al.</i> [56]	TFD, FC		Student (high vs low spatial ability)
Küchemann <i>et al.</i> [57]	TFD	EMA	Progression (predict vs observe phase), performance (high vs low confidence)
Madsen <i>et al.</i> [58]	TFD		Performance (correct vs incorrect)
Madsen <i>et al.</i> [59]	TFD		Stimuli (cues vs no cues), performance (correct vs incorrect), post-hoc progression (eye-movement similarities between trials)
Mason <i>et al.</i> [60]	TFD, FC, TC, other	IA, EMA	Stimuli (reading conditions)
Rouinfar <i>et al.</i> [1]	TFD, MFD	Rubber band model of eye movements and attention	Stimuli (cues vs no cues), progression (initial vs transfer), performance (correct vs incorrect)
Škrabánková <i>et al.</i> [61]	TFD, FC		Student (expertise groups by physics teacher classification)
Smith <i>et al.</i> [62]	TFD, TC		Stimuli (quiz vs homework)
Susac <i>et al.</i> [63]	TFD, FC		Stimuli (with vs without graphical representation)
Susac <i>et al.</i> [64]	TFD		Stimuli (task type), student (physics vs psychology)
Susac <i>et al.</i> [65]	TFD, FC, MFD		Stimuli (with vs without diagram), performance (correct vs incorrect)
Susac <i>et al.</i> [66]	TFD, FC, MFD, other		Performance (correct vs incorrect)
Susac <i>et al.</i> [67]	TFD, FC, other		Student (experimental vs control group), performance (option choice)

(Table continued)

TABLE X. (Continued)

Publication	ET measure ^a	ET theory ^b	Data analysis (data split) ^c
Wu and Liu [68]	TFD, TC		Stimuli (different tasks), student (high vs low prior knowledge)

^aTotal fixation duration (TFD), fixation count (FC), transition count (TC), mean fixation duration (MFD), saccade length (SL), saccadic angle (SA).

^bEye-mind assumption (EMA), immediacy assumption (IA).

^cIndependent variables (cf. Sec. VI).

- [1] A. Rouinfar, E. Agra, A. M. Larson, N. S. Rebello, and L. Loschky, Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams, *Front. Psychol.* **5**, 1094 (2014).
- [2] K. Wright, Eye tracking gets complex, *Physics* **14**, 59 (2021).
- [3] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures* (Oxford University Press, Oxford, 2011).
- [4] A. Duchowski, *Eye Tracking Methodology, Theory and Practice* (Springer, London, 2007).
- [5] A. Gegenfurtner, E. Lehtinen, and R. Säljö, Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains, *Educ. Psychol. Rev.* **23**, 523 (2011).
- [6] M.-L. Lai, M.-J. Tsai, F.-Y. Yang, C.-Y. Hsu, T.-C. Liu, S. W.-Y. Lee, M.-H. Lee, G.-L. Chiou, J.-C. Liang, and C.-C. Tsai, A review of using eye-tracking technology in exploring learning from 2000 to 2012, *Educ. Res. Rev.* **10**, 90 (2013).
- [7] E. Alemdag and K. Cagiltay, A systematic review of eye tracking research on multimedia learning, *Comput. Educ.* **125**, 413 (2018).
- [8] A. R. Strohmaier, K. J. MacKay, A. Obersteiner, and K. M. Reiss, Eye-tracking methodology in mathematics education research: A systematic literature review, *Educ. Stud. Math.* **104**, 147 (2020).
- [9] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *Br. Med. J.* **372**, n71 (2021).
- [10] N. Ariasi and L. Mason, Uncovering the effect of text structure in learning from a science text: An eye-tracking study, *Instr. Sci.* **39**, 581 (2011).
- [11] N. Ariasi, J. Hyönä, J. K. Kaakinen, and L. Mason, An eye-movement analysis of the refutation effect in reading science text, *J. Comput. Assist. Learn.* **33**, 202 (2017).
- [12] G. Trevors, R. Feyzi-Behnagh, R. Azevedo, and F. Bouchet, Self-regulated learning processes vary as a function of epistemic beliefs and contexts: Mixed method evidence from eye tracking and concurrent and retrospective reports, *Learn. Instr.* **42**, 31 (2016).
- [13] Y.-C. Jian, M.-L. Chen, and H.-W. Ko, Context effects in processing of Chinese academic words: An eye-tracking investigation, *Read. Res. Q.* **48**, 403 (2013).
- [14] Y.-C. Jian and H.-W. Ko, Investigating the effects of background knowledge on Chinese word processing during text reading: Evidence from eye movements, *J. Res. Read.* **37**, S71 (2014).
- [15] A. Stott and T. Beelders, The influence of science reading comprehension on South African township learners' learning of science, *S. Afr. J. Sci.* **115**, 1 (2019).
- [16] S. Scrimin and L. Mason, Does mood influence text processing and comprehension? Evidence from an eye-movement study, *Br. J. Educ. Psychol.* **85**, 387 (2015).
- [17] L. Mason, P. Pluchino, and M. C. Tornatora, Effects of picture labeling on science text processing and learning: Evidence from eye movements, *Read. Res. Q.* **48**, 199 (2013).
- [18] P. A. O'Keefe, S. M. Letourneau, B. D. Homer, R. N. Schwartz, and J. L. Plass, Learning from multiple representations: An examination of fixation patterns in a science simulation, *Comput. Hum. Behav.* **35**, 234 (2014).
- [19] M.-J. Tsai, L.-J. Huang, H.-T. Hou, C.-Y. Hsu, and G.-L. Chiou, Visual behavior, flow and achievement in game-based learning, *Comput. Educ.* **98**, 115 (2016).
- [20] M. A. Lindner, A. Eitel, B. Strobel, and O. Köller, Identifying processes underlying the multimedia effect in testing: An eye-movement analysis, *Learn. Instr.* **47**, 91 (2017).
- [21] T. Zu, J. Hutson, L. C. Loschky, and N. S. Rebello, Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment, *J. Educ. Psychol.* **112**, 1338 (2020).
- [22] C.-J. Wu, C.-Y. Liu, C.-H. Yang, and C.-Y. Wu, Children's reading performances in illustrated science texts: Comprehension, eye movements, and interpretation of arrow symbols, *Int. J. Sci. Educ.* **43**, 105 (2021).
- [23] C.-C. Liu, I.-C. Hsieh, C.-T. Wen, M.-H. Chang, S. H. F. Chiang, M.-J. Tsai, C.-J. Chang, and F.-K. Hwang, The affordances and limitations of collaborative science simulations: The analysis from multiple evidences, *Comput. Educ.* **160**, 104029 (2021).

- [24] P.-S. Huang and H.-C. Chen, Gender differences in eye movements in solving text-and-diagram science problems, *Int. J. Sci. Math. Educ.* **14**, 327 (2016).
- [25] M. Andrzejewska and A. Stolińska, Comparing the difficulty of tasks using eye tracking combined with subjective and behavioural criteria, *J. Eye Mov. Res.* **9**, 1 (2016).
- [26] J. Škrabánková, L. Laš, and P. Bujok, Visualisation and interpretation of student strategies in solving natural science-based tasks using the eye-tracker, *J. Eye Mov. Res.* **11**, 1 (2018).
- [27] D. Rosengrant, D. Herrington, and J. O'Brien, Investigating student sustained attention in a guided inquiry lecture course using an eye tracker, *Educ. Psychol. Rev.* **33**, 11 (2021).
- [28] R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [29] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [30] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [31] C. Andrá, P. Lindström, F. Arzarello, K. Holmqvist, O. Robutti, and C. Sabena, Reading mathematics representations: An eye-tracking study, *Int. J. Sci. Math. Educ.* **13**, 237 (2015).
- [32] K. Holmqvist, C. Andrá, P. Lindström, F. Arzarello, F. Ferrara, O. Robutti, and C. Sabena, A method for quantifying focused versus overview behavior in AOI sequences, *Behav. Res.* **43**, 987 (2011).
- [33] R. Groner, F. Walder, and M. Groner, Looking at faces: Local and global aspects of scanpaths, *Adv. Psychology* **22**, 523 (1984).
- [34] H. H. Kornhuber, Blickmotorik, in *Physiologie des Menschen, Band 13: Sehen, Sinnespsychologie*, edited by O. H. Gauer, K. Kramer, and R. Jung (Urban & Schwarzenberg, München, Berlin, Wien, 1978), pp. 357–426.
- [35] R. Brünken, T. Seufert, and S. Zander, Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen, *Z. Pädagogische Psychol.* **19**, 61 (2005).
- [36] M. A. Rau, Conditions for the effectiveness of multiple visual representations in enhancing STEM learning, *Educ. Psychol. Rev.* **29**, 717 (2017).
- [37] Y. Bi and T. Reid, Evaluating students' understanding of statics concepts using eye gaze data, *Int. J. Eng. Educ.* **33**, 225 (2017), <https://sites.ji.sjtu.edu.cn/youyibi/wp-content/uploads/sites/3/2021/03/pdf-3.pdf>.
- [38] S. Brückner, J. Schneider, O. Zlatkin-Troitschanskaia, and H. Drachler, Epistemic network analyses of economics students' graph understanding: An eye-tracking study, *Sensors* **20**, 6908 (2020).
- [39] S. Brückner, O. Zlatkin-Troitschanskaia, S. Küchemann, P. Klein, and J. Kuhn, Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: A postreplication study, *Front. Psychol.* **11**, 2090 (2020).
- [40] S.-C. Chen and H.-C. She, 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 (2020).
- [41] S.-C. Chen, H.-C. She, M.-H. Chuang, J.-Y. Wu, J.-L. Tsai, and T.-P. Jung, Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities, *Comput. Educ.* **74**, 61 (2014).
- [42] K.-P. Chien, C.-Y. Tsai, H.-L. Chen, W.-H. Chang, and S. Chen, Learning differences and eye fixation patterns in virtual and physical science laboratories, *Comput. Educ.* **82**, 191 (2015).
- [43] G.-L. Chiou, C.-Y. Hsu, and M.-J. Tsai, Exploring how students interact with guidance in a physics simulation: Evidence from eye-movement and log data analyses, *Interact. Learn. Environ.*, 1 (2019).
- [44] T. Van Gog, F. Paas, and J. J. G. Van Merriënboer, Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data, *Appl. Cogn. Psychol.* **19**, 205 (2005).
- [45] J. Han, L. Chen, Z. Fu, J. Fritchman, and L. Bao, Eye-tracking of visual attention in web-based assessment using the Force Concept Inventory, *Eur. J. Phys.* **38**, 045702 (2017).
- [46] E. Hejnová and M. Kekule, Observing students problem solving strategies in mechanics by eye-tracking method, *Scientia Educatione* **9**, 102 (2018).
- [47] C. Hoyer and R. Girwidz, 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 (2020).
- [48] B. Ibrahim and L. Ding, Sequential and simultaneous synthesis problem solving: A comparison of students' gaze transitions, *Phys. Rev. Phys. Educ. Res.* **17**, 010126 (2021).
- [49] M. Kekule, Students' approaches when dealing with kinematics graphs explored by eye-tracking research method, *Eur. J. Sci. Math. Educ.* **2**, 108 (2014).
- [50] M. Kekule and J. Viiri, Students' approaches to solving R-FCI tasks observed by eye-tracking method, *Sci. Educat.* **9**, 117 (2018).
- [51] P. Klein, J. Viiri, S. Mozaffari, A. Dengel, and J. Kuhn, Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots?, *Phys. Rev. Phys. Educ. Res.* **14**, 010116 (2018).
- [52] P. Klein, S. Küchemann, S. Brückner, O. Zlatkin-Troitschanskaia, and J. Kuhn, Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students, *Phys. Rev. Phys. Educ. Res.* **15**, 020116 (2019).
- [53] P. Klein, A. Lichtenberger, S. Küchemann, S. Becker, M. Kekule, J. Viiri, C. Baadte, A. Vaterlaus, and J. Kuhn, Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis, *Eur. J. Phys.* **41**, 025701 (2020).
- [54] P. Klein, J. Viiri, and J. Kuhn, Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving, *Phys. Rev. Phys. Educ. Res.* **15**, 010126 (2019).
- [55] P. Klein, S. Becker, S. Küchemann, and J. Kuhn, Test of understanding graphs in kinematics: Item objectives

- confirmed by clustering eye movement transitions, *Phys. Rev. Phys. Educ. Res.* **17**, 013102 (2021).
- [56] M. Kozhevnikov, M. A. Motes, and M. Hegarty, Spatial visualization in physics problem solving, *Cogn. Sci.* **31**, 549 (2007).
- [57] S. Küchemann, P. Klein, H. Fouckhardt, S. Gröber, and J. Kuhn, Students' understanding of non-inertial frames of reference, *Phys. Rev. Phys. Educ. Res.* **16**, 010112 (2020).
- [58] A. M. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Differences in visual attention between those who correctly and incorrectly answer physics problems, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010122 (2012).
- [59] A. Madsen, A. Rouinfar, A. M. Larson, L. C. Loschky, and N. S. Rebello, Can short duration visual cues influence students reasoning and eye movements in physics problems?, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020104 (2013).
- [60] L. Mason, P. Pluchino, M. C. Tornatora, and N. Ariasi, An eye-tracking study of learning from science text with concrete and abstract illustrations, *J. Exp. Educ.* **81**, 356 (2013).
- [61] J. Škrabánková, S. Popelka, and M. Beitlová, Students' ability to work with graphs in physics studies related to three typical student groups, *J. Balt. Sci. Educ.* **19**, 298 (2020).
- [62] A. D. Smith, J. P. Mestre, and B. H. Ross, Eye-gaze patterns as students study worked-out examples in mechanics, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020118 (2010).
- [63] A. Susac, A. Bubic, P. Martinjak, M. Planinic, and M. Palmovic, Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study, *Phys. Rev. Phys. Educ. Res.* **13**, 020125 (2017).
- [64] A. Susac, A. Bubic, E. Kazotti, M. Planinic, and M. Palmovic, Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students, *Phys. Rev. Phys. Educ. Res.* **14**, 020109 (2018).
- [65] A. Susac, A. Bubic, M. Planinic, M. Movre, and M. Palmovic, 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 (2019).
- [66] A. Susac, M. Planinic, A. Bubic, L. Ivanjek, and M. Palmovic, Student recognition of interference and diffraction patterns: An eye-tracking study, *Phys. Rev. Phys. Educ. Res.* **16**, 020133 (2020).
- [67] A. Susac, M. Planinic, A. Bubic, K. Jelcic, L. Ivanjek, K. M. Cvenic, and M. Palmovic, Effect of students' investigative experiments on students' recognition of interference and diffraction patterns: An eye-tracking study, *Phys. Rev. Phys. Educ. Res.* **17**, 010110 (2021).
- [68] C.-J. Wu and C.-Y. Liu, Eye-movement study of high- and low-prior-knowledge students' scientific argumentations with multiple representations, *Phys. Rev. Phys. Educ. Res.* **17**, 010125 (2021).
- [69] E. Agra, D. Johnson, J. Hutson, L. Loschky, and N. S. Rebello, Influence of visual cueing and outcome feedback on students' visual attention during problem solving, in *Proceedings of the 2015 Physics Education Research Conference, College Park, MD*, edited by A. D. Churukian, D. L. Jones, and L. Ding (AIP, New York, 2015), pp. 27–30, [10.1119/perc.2015.pr.002](https://doi.org/10.1119/perc.2015.pr.002).
- [70] R. Holubová, A. Krčmářová, L. Richterek, and J. Říha, Analysis of some selected Force Concept Inventory tasks using eye-tracking and correlation with scientific reasoning skills, in *Proceedings of the 3rd International Baltic Symposium on Science and Technology Education (BalticSTE2019), Šiauliai, Lithuania*, edited by V. Lamanuskas (Scientia Socialis Press, Lithuania, 2019), pp. 63–66.
- [71] M. Kekule, Learning style preferences and their influence on students' problem solving in kinematics observed by eye-tracking method, *AIP Conf. Proc.* **1804**, 030002 (2017).
- [72] P. Klein and J. Kuhn, Analyzing student understanding of vector field plots with respect to divergence, in *Proceedings of the 2017 Physics Education Research Conference, Cincinnati, OH*, edited by L. Ding, A. Traxler, and Y. Cao (AIP, New York, 2017), pp. 220–223, [10.1119/perc.2017.pr.050](https://doi.org/10.1119/perc.2017.pr.050).
- [73] P. Klein, S. Küchemann, P. Van Kampen, L. Doughty, and J. Kuhn, Picture bias in upper-division physics education, in *Frontiers and Advances in Positive Learning in the Age of Information PLATO*, edited by O. Zlatkin-Troitschanskaia (Springer Nature Switzerland, Cham, 2019), pp. 135–142.
- [74] P. Klein, J. Viiri, and J. Kuhn, Visual understanding of divergence and curl: Visual cues promote better learning, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC*, edited by A. Traxler, Y. Cao, and S. Wolf (AIP, New York, 2018), [10.1119/perc.2018.pr.Klein](https://doi.org/10.1119/perc.2018.pr.Klein).
- [75] S. Küchemann, P. Klein, S. Becker, N. Kumari, and J. Kuhn, Classification of students' conceptual understanding in STEM education using their visual attention distributions: A comparison of three machine-learning approaches, in *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU '20)* edited by H. C. Lane, S. Zvacek, and J. Uhomobhi (SciTePress, Setbal, Portugal, 2020), [10.5220/0009359400360046](https://doi.org/10.5220/0009359400360046).
- [76] A. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Using ScanMatch scores to understand differences in eye movements between correct and incorrect solvers on physics problems, in *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12), Santa Barbara, California* (Association for Computing Machinery, New York, 2012), pp. 193–196.
- [77] A. M. Madsen, A. Rouinfar, A. M. Larson, L. C. Loschky, and N. S. Rebello, Do perceptually salient elements in physics problems influence students' eye movements and answers?, *AIP Conf. Proc.* **1513**, 274 (2013).
- [78] S. Mozaffari, P. Klein, J. Viiri, S. Ahmed, J. Kuhn, and A. Dengel, Evaluating similarity measures for gaze patterns in the context of representational competence in physics education, in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18), Warsaw, Poland* (Association for Computing Machinery, New York, 2018), pp. 1–5.
- [79] S. Mozaffari, M. Al-Naser, P. Klein, S. Küchemann, J. Kuhn, T. Widmann, and A. Dengel, Classification of visual strategies in physics vector field problem solving, in *Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART '20), Valletta*,

- Malta*, edited by A. Rocha, L. Steels, and J. Van den Herik (SciTePress, Setbal, Portugal, 2020), pp. 257–267, [10.5220/0009173902570267](https://doi.org/10.5220/0009173902570267).
- [80] S. Mozaffari, P. Klein, M. Al-Naser, S. Küchemann, J. Kuhn, T. Widmann, and A. Dengel, Quantifying gaze-based strategic patterns in physics vector field divergence, in *Agents and Artificial Intelligence, ICAART '20. Lecture Notes in Computer Science*, Vol. 12613, edited by A. P. Rocha, L. Steels, and J. Van den Herik (Springer International Publishing, Basel, 2021), pp. 465–481.
- [81] S. Mozaffari, J. Kuhn, P. Klein, A. Dengel, and S. S. Bukhari, Entropy based transition analysis of eye movement on physics representational competence, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), Heidelberg, Germany* (Association for Computing Machinery, New York, 2016), pp. 1027–1034, [10.1145/2968219.2968564](https://doi.org/10.1145/2968219.2968564).
- [82] E. Ohno, A. Shimojo, and M. Iwata, Analysis of problem solving processes in physics based on eye-movement data, in *Proceedings of GIREP-EPEC 2015 International Conference, Wroclaw, Poland, (Key Competences in Physics Teaching and Learning)*, edited by T. Greczylo and E. Debowska (Institute of Experimental Physics, University of Wroclaw, Wroclaw, Poland, 2015), pp. 64–70.
- [83] D. Rosengrant, C. Thomson, and T. Mzoughi, Comparing experts and novices in solving electrical circuit problems with the help of eye-tracking, *AIP Conf. Proc.* **1179**, 249 (2009).
- [84] D. Rosengrant, Gaze scribing in physics problem solving, in *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '10), Austin, Texas* (Association for Computing Machinery, New York, 2010), pp. 45–48.
- [85] R. Rosiek, M. Sajka, E. Ohno, A. Shimojo, M. Iwata, and D. Wcislo, An excerpt from an eye-tracking comparative study between Poland and Japan with the use of Force Concept Inventory, *AIP Conf. Proc.* **1804**, 060003 (2017).
- [86] R. Rosiek and M. Sajka, One task - Many strategies of interpreting and reasons of decision making in the context of an eye-tracking research, *AIP Conf. Proc.* **2152**, 030028 (2019).
- [87] A. Rouinfar, E. Agra, J. Murray, A. M. Larson, L. C. Loschky, and N. S. Rebello, Can visual cues and correctness feedback influence students reasoning?, in *Proceedings of the 2013 Physics Education Research Conference, Portland, OR*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (AIP, New York, 2013), pp. 305–308, [10.1119/perc.2013.pr.064](https://doi.org/10.1119/perc.2013.pr.064).
- [88] A. Rouinfar, E. Agra, J. Murray, A. M. Larson, and L. C. Loschky, Influence of visual cueing on students' eye movements while solving physics problems, *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14), Safety Harbor, Florida* (Association for Computing Machinery, New York, 2014), pp. 191–194.
- [89] J. Viiri, M. Kekule, J. Isoniemi, and J. Hautala, Eye-tracking the effects of representation on students problem solving approaches, in *Proceedings of the Finnish Mathematics and Science Education Research Association (FMSERA) Annual Symposium, 2017*, edited by M. A. Asikainen and P. E. Hirvonen (FMSERA Journal, Helsinki, Finland, 2017), pp. 88–98, <https://journal.fi/fmsera/article/view/60941/27043>.
- [90] J. Viiri, J. Kilpeläinen, M. Kekule, E. Ohno, and J. Hautala, Eye-movement study of mechanics problem solving using multimodal options, in *Research and Innovation in Physics Education: Two Sides of the Same Coin*, edited by J. Guisasaola and K. Zuza (Springer Nature Switzerland, Cham, 2020), pp. 145–154.
- [91] X. Wu, J. Hutson, L. C. Loschky, and N. S. Rebello, How do multimedia hints affect students eye movements on conceptual physics problems?, in *Proceedings of the 2015 Physics Education Research Conference, College Park, MD*, edited by A. D. Churukian, D. L. Jones, and L. Ding (AIP, New York, 2015), pp. 383–386, [10.1119/perc.2015.pr.091](https://doi.org/10.1119/perc.2015.pr.091).
- [92] E. Agra, A conceptual model for facilitating learning from physics tasks using visual cueing and outcome feedback: Theory and experiments, Ph.D. thesis, Kansas State University, Manhattan, Kansas, 2015.
- [93] X. Wu, Influence of multimedia hints on conceptual physics problem solving and visual attention, Ph.D. thesis, Kansas State University, Manhattan, Kansas, 2016.
- [94] A. Rouinfar, Influence of visual cueing and outcome feedback on physics problem solving and visual attention, Ph.D. thesis, Kansas State University, Manhattan, Kansas, 2014.
- [95] A. M. Madsen, Studies of visual attention in physics problem solving, Ph.D. thesis, Kansas State University, Manhattan, Kansas, 2013.
- [96] D. D. Salvucci and J. H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '00), Palm Beach Gardens, Florida* (Association for Computing Machinery, New York, 2000), pp. 71–78.
- [97] M. A. Just and P. Carpenter, A theory of reading: From eye fixations to comprehension, *Psychol. Rev.* **87**, 329 (1980).
- [98] J. M. Henderson, Visual attention and eye movement control during reading and picture viewing, in *Eye Movements and Visual Cognition*, edited by K. Rayner (Springer Series in Neuropsychology, New York, 1992), pp. 260–283, [10.1007/978-1-4612-2852-3_15](https://doi.org/10.1007/978-1-4612-2852-3_15).
- [99] J. M. Henderson, Visual attention and saccadic eye movements, in *Perception and Cognition: Advances in Eye Movement Research. Studies in Visual Information Processing*, edited by G. d'Ydewalle and J. Van Rensbergen (North-Holland/Elsevier Science Publishers, Amsterdam, 1993), pp. 37–50.
- [100] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden, Quantitative analysis of culture using millions of digitized books, *Science* **331**, 176 (2011).
- [101] Google Ngram Viewer, https://books.google.com/ngrams/graph?content=eye+tracking&year_start=2014&year_end=2019&corpus=26&smoothing=3.
- [102] PhysPort—Supporting physics teaching with research-based resources, <https://www.physport.org>.

- [103] L. Bao and K. Koenig, Physics education research for 21st century learning, *Discip. Interdiscip. Sci. Educ. Res.* **1**, 2 (2019).
- [104] H. Schmidt, B. Wassermann, and G. Zimmermann, An adaptive and adaptable learning platform with realtime eye-tracking support: lessons learned, in *DeLFI 2014—Die 12. e-Learning Fachtagung Informatik*, edited by S. Trahasch, R. Plötzner, G. Schneider, D. Sassi, C. Gayer, and N. Wöhrle (Gesellschaft für Informatik e.V., Bonn, 2014), pp. 241–252.
- [105] H. Jarodzka, H. Gruber, and K. Holmqvist, Eye tracking in educational science: Theoretical frameworks and research agendas, *J. Eye Mov. Res.* **10**, 1 (2017).
- [106] A. Nehring and S. Busch, Chemistry demonstrations and visual attention: Does the setup matter? Evidence from a double-blinded eye-tracking study, *J. Chem. Educ.* **95**, 1724 (2018).
- [107] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. Pea, Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups, *Int. J. Comput. Support. Collab. Learn.* **13**, 241 (2018).
- [108] P. Klein, A. Müller, and J. Kuhn, Assessment of representational competence in kinematics, *Phys. Rev. Phys. Educ. Res.* **13**, 010132 (2017).
- [109] A. Kohnle, S. E. Ainsworth, and G. Passante, Sketching to support visual learning with interactive tutorials, *Phys. Rev. Phys. Educ. Res.* **16**, 020139 (2020).
- [110] S. B. McKagan, K. K. Perkins, M. Dubson, C. Malley, S. Reid, R. LeMaster, and C. E. Wieman, Developing and researching PhET simulations for teaching quantum mechanics, *Am. J. Phys.* **76**, 406 (2008).
- [111] C. Singh, M. Belloni, and W. Christian, Improving students understanding of quantum mechanics, *Phys. Today* **59**, No. 8, 43 (2006).
- [112] T. Van Gog and K. Scheiter, Eye tracking as a tool to study and enhance multimedia learning, *Learn. Instr.* **20**, 95 (2010).
- [113] K. L. Havanki and J. R. VandenPlas, Eye tracking methodology for chemistry education research, in *Tools of Chemistry Education Research*, edited by D. M. Bunce and R. S. Cole (ACS Publications, Washington DC, 2014), pp. 191–218, [10.1021/bk-2014-1166.ch011](https://doi.org/10.1021/bk-2014-1166.ch011).
- [114] R. J. Marzano, D. J. Pickering, and J. E. Pollock, *Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement* (Association for Supervision and Curriculum Development, Alexandria, VA, 2001).
- [115] M. A. Ruiz-Primo, D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard, Impact of undergraduate science course innovations on learning, *Science* **331**, 1269 (2011).
- [116] C. Wieman, Why not try a scientific approach to science education?, *Change* **39**, 9 (2007).
- [117] E. M. Kok and H. Jarodzka, Before your very eyes: The value and limitations of eye tracking in medical education, *Med. Educ.* **51**, 114 (2017).
- [118] S. Cullipher, S. J. Hansen, and J. R. VandenPlas, Eye tracking as a research tool: An introduction, in *In Eye Tracking for the Chemistry Education Researcher*, edited by J. R. VandenPlas, S. J. R. Hansen, and S. Cullipher (American Chemical Society, Washington, DC, 2018), pp. 1–9, [10.1021/bk-2018-1292.ch001](https://doi.org/10.1021/bk-2018-1292.ch001).
- [119] J. Hyönä, The use of eye movements in the study of multimedia learning, *Learn. Instr.* **20**, 172 (2010).
- [120] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters, Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts, *Surg. Endosc.* **24**, 2458 (2010).
- [121] H. Ashraf, M. H. Sodergren, N. Merali, G. Mylonas, H. Singh, and A. Darzi, Eye-tracking technology in medical education: A systematic review, *Med. Teach.* **40**, 62 (2018).
- [122] K. Krejtz, A. T. Duchowski, I. Krejtz, A. Kopacz, and P. Chrzastowski-Wachtel, Gaze transitions when learning with multimedia, *J. Eye. Mov. Res.* **9**, 1 (2016).
- [123] J. Holsanova, N. Holmberg, and K. Holmqvist, Reading information graphics: The role of spatial contiguity and dual attentional guidance, *Appl. Cogn. Psychol.* **23**, 1215 (2009).
- [124] A. Schüler, Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration, *Learn. Instr.* **49**, 218 (2017).
- [125] R. E. Mayer, Cognitive theory of multimedia learning, in *The Cambridge Handbook of Multimedia Learning* (Cambridge University Press, New York, 2005), pp. 31–48.
- [126] A. R. Gardner-Medwin, Confidence assessment in the teaching of basic science, *ALT-J* **3**, 80 (1995).
- [127] J. E. Dowd, I. Araujo, and E. Mazur, Making sense of confusion: Relating performance, confidence, and self-efficacy to expressions of confusion in an introductory physics class, *Phys. Rev. ST Phys. Educ. Res.* **11**, 010107 (2015).
- [128] J. Scheid, A. Müller, R. Hettmannsperger, and W. Schnotz, Improving learners representational coherence ability with experiment-related representational activity tasks, *Phys. Rev. Phys. Educ. Res.* **15**, 010142 (2019).
- [129] M. Opfermann, A. Schmeck, and H. E. Fischer, Multiple representations in physics and science education: Why should we use them?, in *Multiple Representations in Physics Education*, edited by D. Treagust, R. Duit, and H. Fischer (Springer, Cham, 2017), pp. 1–22, [10.1007/978-3-319-58914-5_1](https://doi.org/10.1007/978-3-319-58914-5_1).
- [130] M. Prokop, L. Pilař, and I. Tichá, Impact of think-aloud on eye-tracking: A comparison of concurrent and retrospective think-aloud for research on decision-making in the game environment, *Sensors* **20**, 2750 (2020).
- [131] T. Van Gog, F. Paas, J. J. Van Merriënboer, and P. Witte, Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting, *J. Exp. Psychol. Gen.* **11**, 237 (2005).
- [132] B. Schneider and R. Pea, Real-time mutual gaze perception enhances collaborative learning and collaboration quality, *Intern. J. Comput.-Support. Collab. Learn.* **8**, 375 (2013).