

Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students

P. Klein,^{1,*} S. Küchemann,¹ S. Brückner,² O. Zlatkin-Troitschanskaia,² and J. Kuhn¹

¹*Department of Physics, Physics Education Research Group, Technische Universität Kaiserslautern, Erwin-Schrödinger-Straße 46, 67663 Kaiserslautern, Germany*

²*Faculty of Law, Management and Economics, Department of Business and Economics Education, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 9, 55128 Mainz, Germany*



(Received 22 January 2019; published 14 August 2019)

The understanding of graphs and extraction of relevant information from graphs plays a major role in physics education and is also important in several related fields. Recently, Susac *et al.* [*Phys. Rev. Phys. Educ. Res.* **14**, 020109 (2018)] compared physics and psychology students' understanding of graphs in the contexts of physics and finance. They showed that physicists scored significantly higher in both domains and that all students solved the slope problems better than the area problems. Moreover, eye-tracking data revealed that physics students spent more time on problems associated with the area under the graph and focused longer on the axis labels of finance graphs, indicating higher cognitive demands. In this eye-tracking study, we aim for a generalization of the results obtained by Susac *et al.* by comparing physics students to another nonphysics sample, viz., economics students. The findings broadly confirm the results of Susac *et al.*; that is, physics students perform better than nonphysics students. While economics students likely have better prior knowledge on finance context than psychology students, the physics students still outperform them on the finance questions. In contrast to the work by Susac *et al.*, both groups of students had the same visit duration on the graphs, consequently proving total dwell time to be an inadequate predictor of performance. Instead, we identify that attention on concept-specific areas of interest within the graphs discriminates the correct from the incorrect performers. Furthermore, we analyzed the confidence level of the two student groups and found that physics students have a higher ability to correctly judge their own performance compared to economics students. Overall, our results highlight the importance of an instructional adjustment towards a more mathematical- and graphical-based education.

DOI: [10.1103/PhysRevPhysEducRes.15.020116](https://doi.org/10.1103/PhysRevPhysEducRes.15.020116)

I. INTRODUCTION

For several years, students' understanding of graphs has stimulated major research efforts in physics and other fields [1–4]. In physics, these studies have identified several difficulties for students, such as the point-interval confusion, variable confusion, or the area-slope-height confusion [1].

Going beyond the study of mathematical prerequisites concerning the slope and the area under the graph, some studies focus on the influence of context in which the graph is presented and the question is posed [5]. In such a case, besides the mathematical skills for determining the area and the slope of a graph, it is necessary to identify the connection between the desired quantity and

the information displayed in the graph. For example, in the context of physics, it is necessary to know that acceleration can be achieved by determining the slope in a velocity-time graph. This so-called domain knowledge develops with the level of exposure to the content. When a problem is posed in a different context, the student needs to transfer the mathematical skills to the different domain. Previous research has shown that there are two main reasons why this transfer could fail: either the student does not possess the required mathematical skill or the corresponding mental resource does exist, but it is not activated due to a wrong interpretation of the presented problem [6].

Recently, Susac *et al.* compared the qualitative and quantitative graphical understanding of Croatian physics students and psychology students and found that both groups struggle more with questions about the concept of an area under the curve than with problems about the slope [7]. With four isomorphic pairs of questions which required the same mathematical procedure in the context of physics and finance, they showed that physics students focused significantly longer on the graph and spent more

*pklein@physik.uni-kl.de

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

time looking at questions in the unfamiliar context of finance. The high scores of physics students in comparison to psychology students in the finance problems indicate that the mathematical skills of physics students are an important asset for the transfer of knowledge to unfamiliar contexts. As a consequence, Susac *et al.* [7] suggest emphasizing the concept of the area under a graph and the interpretation of graphs during instructions since they are important in manifold contexts.

Besides the evaluation of standard test scores, the rating of confidence levels with multiple-choice questions recently became popular in physics education research (PER) [8–10]. From studies with economics and psychology students, it is known that high-performing students have a better ability to accurately judge their own performance [11]. This observation can be quantified by comparing the confidence ratings associated with correct responses with those of incorrect responses [12]. The more accurate the judgment, the higher the difference between the scores of correct and incorrect responses. Moreover, it was found that poorly performing students grossly overestimated their performance, which is widely known as the Dunning-Kruger effect [13]. This effect was recently validated in PER [9] and could also be observed in problems testing the representational competence of physics students in the context of kinematics [14]. However, little is known about the influence of the subject context (e.g., finance or physics) on students' confidence ratings and about the ability to judge one's own performance in the context of graphs. Since confidence ratings are found to be an efficient tool to indicate the existence of misconceptions and learning difficulties [14], the assessment of students' response confidence rating in this study enriches the research about students' misconceptions [8].

To examine students' response processes during graphical task solving, eye tracking has been proven to be a powerful tool to complement the research with a data resource on students' visual attention [15]. Previous eye-tracking research on students' understanding of kinematic graphs showed a strong link between the visuospatial abilities of a student and the understanding of kinematic graphs, which is linked to the tendency of low-spatial students to interpret graphs literally [16]. In this context, Madsen *et al.* showed that students who answered a question correctly focus longer on relevant areas of a graph such as the axes [17]. Their finding also suggests that previous exposure to the context of a problem enhances the focus on the important regions [18]. Consequently, the abovementioned learning difficulties and misconceptions, e.g., the point-interval confusion, are likely to be reflected in certain eye-movement patterns and attentional distributions shifted towards conceptually irrelevant regions. In fact, Kekule reported qualitative results that indicate different task-solving approaches between best- and worst-performing students when they solved the test of

understanding graphs in kinematics (TUG-K) [19,20]. On the item level, it can be expected that different cognitive demands (e.g., verbal vs quantitative reasoning) also create different eye-movement patterns. However, prior research has mainly investigated total dwell time on different parts of the multiple-choice items (question statement, diagrams, and alternatives), leaving this assumption unproven. In this context, Kekule found no differences in the average fixation duration between best- and worst-performing students while solving tasks of the TUG-K [20]. The conclusions of eye-tracking studies have the potential to identify misconceptions and learning difficulties in order to design particular instructions to help students learning.

In this paper, we present a replication study of the original work of Susac *et al.* [7] in which we compare the performance of physics and economics students using the identical questions from the original work about graph slope and area under the graph in the context of physics and finance. To generalize the body of evidence, we compare physics students to a different group of nonphysics students, viz., economics students instead of psychology students. The study presented here can be considered as an operational replication within the classification of Lykken [21] or partial direct replication according to Schmidt [22] because we use the main test material of the original work without the inclusion of certain control variables and we extend the sample group since we intend to test the generalizability of the influence of context on the understanding of kinematic graphs. With this modification in mind we are unable to test whether or not the original findings are correct, but we are able to test the veracity of the conclusion and constrain the generalizability of certain findings in the original manuscript, as discussed below. While we acknowledge the value of a direct replication in the sense of Schmidt, we found it important to extend the previously tested target group to students with a potentially different access to mathematical methods and problems in the context of finance due to their chosen major in economics.

In general, a replication study constitutes an essential aspect of empirical research which serves to corroborate and constrain previous findings and, thus, emphasizes and, ideally, accelerates the application of important aspects in teaching practices. Successful examples of such important replication studies can be found in several areas of physics education research [23–25]. Furthermore, replications inhibit a particular value when an independent group which did not invent the examined theory or protocol aims to reproduce the original results [26], since it allows one to test for certain biases, as pointed out by Makel and Plucker [27], or experimental design errors. They inhibit value for educational research because they empower the empirical body of research in terms of generality and independence of particular samples. The motivation for this replication study is to understand and consider the added complexity when

teaching kinematic graphs or general visual representations in the field of physics as well as in other fields. We did not presume or anticipate any biases or weaknesses in the experimental design in the original study by Susac *et al.* and we also did not find an indication for any shortcoming. Eye tracking is used here as an important tool to identify correct task-solving approaches—even in different contexts.

In this replication study, we consolidate and constrain the results of the original work of Susac *et al.* In addition, we address three main research questions.

- (1) How do the physics students and the economics students solve tasks associated with slopes and the area under a graph in the context of physics and in the context of finance?
- (2) How do physics students rate their confidence when giving a correct or false answer in comparison to economics students?
- (3) Do the eye movements of the students in each group reveal differences with respect to their performance?

We are particularly interested in the pieces of information that the different student groups allocate their attention to. Therefore, we apply deeper analysis procedures to the eye-tracking data (see Sec. II E), allowing us to investigate how students visually interact with the diagrams.

II. METHODS

A. Participants

A total of 69 students from two German universities took part in the experiment. Twenty-nine ($N = 29$) of the participants were first-year students from the Technische Universität Kaiserslautern majoring in physics and the other participants ($N = 40$) were economics students from the University of Mainz who had completed at least one course on finance. The age of the physics students ranged from 18 to 27 years; their average age was 19.9 ± 1.8 years. The age of the economics students ranged from 18 to 31 years; their average age was 21.6 ± 2.7 years. Participation was voluntary, took about 45 minutes, and was compensated with 10€. All students had normal or correct-to-normal vision.

Physics is taught as a compulsory subject for four years in the lower secondary level (grade 5–10) in Germany. In kinematics, the basic quantities (time, displacement) are introduced and basic ideas are taught, such as the concept of velocity and acceleration and their (indirect) measurement. Explicit teaching of kinematics graphs starts in physics courses at the beginning of the upper secondary level (participant age 15–16 years), where physics is not a compulsory subject. Most of the physics students (89%) selected the physics course at school whereas only a minority of economics students (16%) did so. None of the participants were exposed to kinematics graphs after school. The physics students, in particular, had not learned about kinematic graphs in university courses since the

experiment took place during the first weeks of the students' first academic year. According to faculty, the economics students also did not encounter graphs similar to those used in this study in their finance courses. Obviously, economics students did not learn about kinematics graphs after school since kinematics is not part of an economics curriculum.

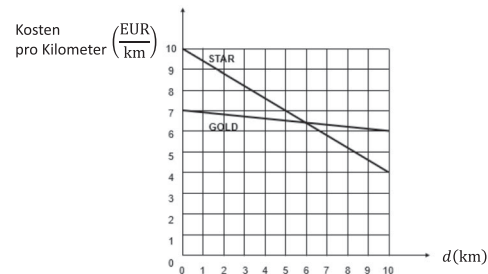
B. Materials

In this study, we used the same eight multiple-choice items from the study of Susac *et al.* [7] and translated them into German [28] (see Fig. 1 for an example). All items showed linear graphs and tested student's understanding of graph slope and the area under a graph with the help of qualitative and quantitative questions. Each combination of concepts (graph slope, area under graph) and question type (qualitative or quantitative) was presented in the context of physics and finance, yielding four sets of isomorphic questions.

C. Apparatus

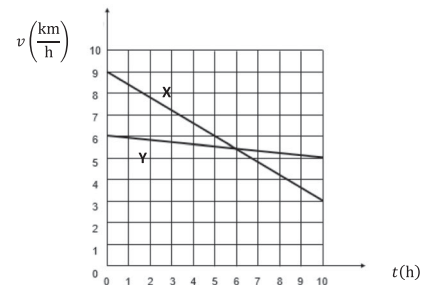
The tasks were presented on a 22-in. computer screen. The resolution of the computer screen was set to

Das Diagramm zeigt den Zusammenhang zwischen den Kosten für einen zurückgelegten Kilometer und der Fahrdistanz für die zwei Taxiunternehmen „Star“ und „Gold“. Vergleichen Sie die Kosten, die für eine 8 km-lange Taxifahrt anfallen.



- (a) Die Kosten des Unternehmens „Gold“ sind geringer als die Kosten des Unternehmens „Star“.
- (b) Die Kosten des Unternehmens „Gold“ sind genauso groß wie die Kosten des Unternehmens „Star“.
- (c) Die Kosten des Unternehmens „Gold“ sind größer als die Kosten des Unternehmens „Star“.

Das Diagramm zeigt den zeitlichen Verlauf der Geschwindigkeiten zweier Objekte X und Y. Vergleichen Sie den zurückgelegten Weg beider Objekte während den ersten 9 Stunden.



- (a) Der von Objekt X zurückgelegte Weg ist kürzer als der von Objekt Y zurückgelegte Weg.
- (b) Der von Objekt X zurückgelegte Weg ist genauso groß wie der von Objekt Y zurückgelegte Weg.
- (c) Der von Objekt X zurückgelegte Weg ist länger als der von Objekt Y zurückgelegte Weg.

FIG. 1. An isomorphic pair of qualitative questions about the area under a curve.

1920 × 1080 pixels with a refresh rate of 75 Hz. Eye movements were recorded with a Tobii X3-120 stationary eye-tracking system [29], which had an accuracy of less than 0.40° of visual angle (as reported by the manufacturer) and a sampling frequency of 120 Hz. The system allows a relatively high degree of freedom in terms of head movement (no chin rest was used). To detect fixations and saccades, an I-VT algorithm was adopted [30]. An eye movement was classified as a saccade (i.e., in motion) if the acceleration of the eyes exceeded 8500°/s² and velocity exceeded 30°/s.

D. Procedure and measures

First, the students were introduced to the eye tracker and a 9-point calibration process was used for a fully customized and accurate gaze point calculation. Then, the eight multiple-choice items were presented in a partially counter-balanced sequence (i.e., isomorphic questions were never presented one after another). Each slide contained the question, the diagram, and the answer options. After students were ready to give their answer, they pressed a key to advance to the next page, where they entered their answer and their response confidence on a 6-point Likert-type rating scale, ranging from very high confidence to very low confidence. For the latter, the students were asked how confident they were about the correctness of the answer they had just given. The students could take as much time as needed to answer a question. After the eye movements had been measured, students answered a short questionnaire about their demographics.

In contrast to Susac *et al.*, we did not assign a paper-and-pencil test asking for written explanations of students' answers since their choice of strategy would have exceeded the scope of this research. However, we measured students' response confidence to obtain information about correct answers selected by construct-irrelevant response behavior (e.g., random guessing).

E. Data analysis

Each correct answer during the eye-tracking measurement was credited with 1 point and a false answer received 0 points, yielding a maximum of 8 points. With the absence of given explanations, no correction to the scores was made. We found that the combination of correct answers and very low response confidence (lucky guess, cf. Ref. [8]) was very rare (less than 3% of all cases). The confidence ratings were linearly transformed to a [0,1] scale, where 0 means lowest and 1 means highest confidence.

Part of the analysis procedure was adopted from Susac *et al.*, i.e., areas of interest (AOIs) were defined that covered different parts of the items. The total dwell time was calculated from the eye-tracking data related to the introduction text and question, the graph, the axis labels, the axis tick labels, and the multiple-choice answers; see Fig. 2(a). The AOIs *Question*, *Graph* and *Multiple choice* did not

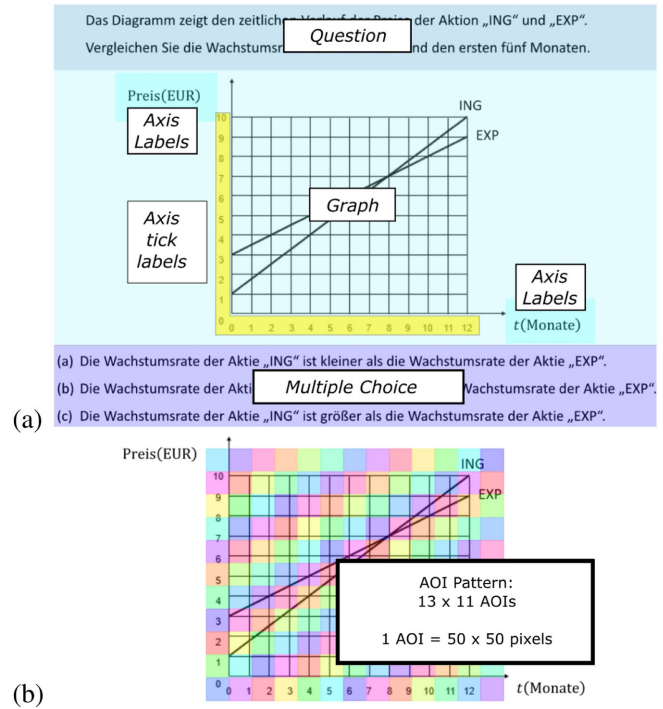


FIG. 2. (a) Definition of AOIs for one item. (b) AOI pattern covering the diagram (without axis labels) for the same item.

overlap and covered the full presentation slide whereas the AOIs *Axis labels* and *Axis tick labels* were embedded into the *Graph* AOI. Moreover, we investigated whether the spatial distribution of visual attention differs between physics and economics students. For this purpose, the diagram of each item was covered with an array of small equally sized AOIs $\{G_1, \dots, G_n\}$ (AOI size 50 × 50 pixels) and the students' fixation count was determined for each AOI G_i [31]. Average fixation counts were computed within each group and normalized scores were compared across groups to identify spots that attract more attention from either physics or economics students.

While fixation data provide information about where students pay visual attention, the absolute saccadic direction contains complementary information about the direction in which students were looking at the line graphs. The absolute saccadic direction is the angle of any saccadic movement with respect to a horizontal line (0°). The absolute saccadic direction was calculated by applying basic trigonometry on two consecutive fixations (Ref. [32], p. 303) based on the fixation data on the AOI *Graph*. To further analyze the saccade data quantitatively, the saccades that corresponded to the gradient angles of the line graph (within a 5° tolerance band) were counted for each student. By this means, it can be concluded to what extent students followed the line graphs with their eyes.

To determine the effects of question type, concepts, group, and context, several analyses of variance (ANOVAs) were conducted. All datasets that were analyzed satisfied

TABLE I. Comparison of methodological features between the study presented here and the original study from Susac *et al.* [7].

Method	This study	Original study [7]
Participants	27 physics students (first year), 40 economics students	45 physics students (teacher program, fourth year), 45 psychology students
Materials	Four isomorphic pairs of items about graph slope and area under a curve in the context of physics and finance	
Apparatus	tobii X3–120 Hz	SMI RED500 Hz
Additional data	Confidence scores	Student strategies (questionnaire)
Coding scheme	Only direct response (correct or incorrect)	Response and explanation (correction)
Data analysis	ANOVAs to determine the effects of question type, concepts, group, and context on the dependent variables AOIs question, graph, multiple choice, axis labels AOI axis tick labels, saccadic direction, attention distribution	Analysis of student strategies

the assumptions required for conducting ANOVAs. A threshold of $p = 0.05$ was used for determining the level of effect significance within all conducted tests. The Levene test was used to test the null hypothesis that the variance is equal across groups. The assumption of homogeneity of variance was met for every comparison. The eye-tracking data of two physics students were invalid; thus, those two participants were removed from all analyses. When reporting correlation coefficients, the Pearson correlation coefficient was used.

F. Comparison to the study of Susac *et al.*

In the Introduction, the motivation for carrying out this study was formulated and the importance of replication studies in education research was emphasized. In Table I, we briefly summarize the major methodological differences between the study presented here and the original study from Susac *et al.* [7].

The physics sample tested by Susac *et al.* is substantially different from the sample tested here due to their different target degree (physics majors vs prospective physics teachers) and their study progress (first year vs fourth year). On the one hand, the difference between both physics samples expands the data basis towards more generalization of the research question in terms of comparing “physics” vs nonphysics students. On the other hand, a direct comparison between the results obtained in both studies must consider these differences. Apart from that, Susac *et al.* assessed students’ strategies in solving problems with graphs and used this information to adjust the performance scores. Instead, we assessed the students’ confidence scores to obtain additional information about their (metacognitive skills), and thus used raw test scores for the analysis. As a consequence, the performance scores cannot be compared directly between both studies. However, the measured effects can be compared since the same data analysis approach was chosen.

For the eye-tracking data, the same areas of interest were defined as in the original work and the same analysis procedure was chosen. Because of the same methodology,

the gaze data measured in this study can directly be compared to the gaze data measured by Susac *et al.*, and they are presented next to each other, allowing for a convenient comparison between the two. Extending previous methods, the spatial distribution of visual attention was investigated on the item level and the saccadic directions were analyzed to obtain more information about the conceptual and perceptual approaches when students solve graph problems.

III. RESULTS

In this section, the analysis of students’ scores, response confidence ratings, and eye-tracking data is reported. Descriptive data in figures are presented together with data from the original work to enable a convenient comparison across both studies.

A. Analysis of students’ scores

The mean test score and standard deviation were $(55 \pm 27)\%$. The physics students $[(69 \pm 26)\%]$ scored significantly higher than the economics students $[(46 \pm 23)\%]$, $p = 2 \times 10^{-4}$.

To evaluate the impact of question type (qualitative vs quantitative) and concept (graph slope vs area under a graph) on the students’ scores, a two-way repeated-measure ANOVA was conducted for each group of students. For physics students, we found a statistically significant main effect of both factors, type of question [$F(1, 26) = 7.27$, $p < 0.01$, $\eta_p^2 = 0.22$] and concept [$F(1, 26) = 4.32$, $p < 0.05$, $\eta_p^2 = 0.14$], whereas the interaction effect was not significant [$F(1, 26) = 0.13$, $p > 0.05$, $\eta_p^2 = 0.01$]. For economics students, we also found a statistically significant main effect of both factors, type of question [$F(1, 39) = 49.9$, $p < 0.001$, $\eta_p^2 = 0.56$] and concept [$F(1, 39) = 21.5$, $p < 0.001$, $\eta_p^2 = 0.36$]. Moreover, the interaction was significant as well [$F(1, 39) = 8.39$, $p < 0.01$, $\eta_p^2 = 0.17$]. Both groups of students solved qualitative questions better than quantitative questions and had higher scores on questions about graph slope than

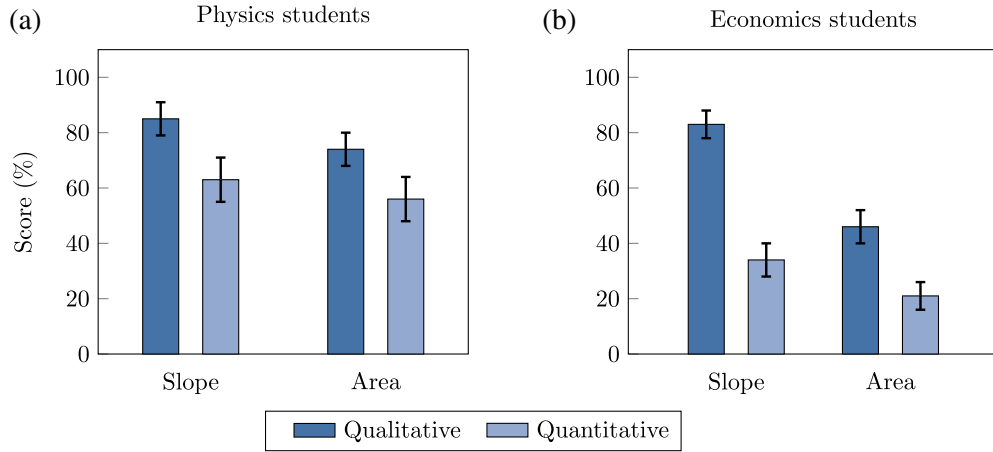


FIG. 3. Average scores of (a) physics students and (b) economics students on the qualitative and quantitative questions about graph slope and the area under a graph. The error bars represent 1 standard error of the mean (SEM).

on questions about the area under a graph (Fig. 3). For the economics students, the difference between scores on qualitative and quantitative questions was larger for questions about graph slope [Fig. 3(b)].

To compare both groups of students across both contexts, we applied a repeated-measure ANOVA with the context (physics vs finance) as a within-subject factor and with the group as a between-subject factor. The analysis was performed for each pair of isomorphic questions, and the results are presented in Table II. For qualitative questions about the slope concept, there were no group or context effects and no interaction effects. For quantitative questions on graph slope, physics students achieved higher scores than economics students, and both groups of students scored higher on the physics questions than on the finance questions, yielding no interaction effect (Fig. 4). For qualitative questions on the area under a graph, physics students had higher scores than economics students. Both physics and economics students solved the physics questions better than the finance questions; hence no interaction effect occurred. For quantitative questions on the area under a graph, physics students scored higher than economics students did. Overall, there was no difference between the scores on physics and finance questions. There was a marginal interaction effect between student group and context, and further data analysis revealed that economics students scored higher on the finance questions than on

physics questions [$t(39) = 2.01$, $p = 0.06$]. There was an opposite trend for the physics students, but the difference was not significant.

B. Analysis of students' confidence ratings

The mean confidence rating and standard deviation were $(62 \pm 23)\%$. The physics students reported a confidence level of $(66 \pm 27)\%$ and the economics students reported a confidence level of $(59 \pm 19)\%$, with no significant differences between the groups, $p > 0.05$. For the physics students, the total score and the mean confidence level were highly correlated [$r(27) = 0.51$, $p < 0.01$], whereas for the economics students, there was no significant correlation [$r(40) = 0.28$, $p > 0.05$].

To further explore students' confidence ratings, we applied the same analysis procedure that we applied to the scores. Two-way ANOVAs revealed no significant main effects of the factors concept and type of question for either group of students. However, the interaction between question type and concept was significant for physics students [$F(1, 26) = 4.75$, $p < 0.05$, $\eta_p^2 = 0.15$] and more pronounced for economics students [$F(1, 39) = 10.67$, $p < 0.01$, $\eta_p^2 = 0.22$]. For questions about graph slope, students are more confident about their responses to qualitative questions than to quantitative questions, and the opposite applies to the area questions (Fig. 12 in the

TABLE II. Results of the two-way ANOVAs conducted on students' scores with the context (physics vs finance) as a within-subject factor and with the group (physics students vs economics students) as a between-subject factor.

	Group			Context			Interaction		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Slope qualitative	0.12	>0.05	...	0.49	>0.05	...	0.49	>0.05	...
Slope quantitative	9.42	0.003	0.13	13.9	$<10^{-4}$	0.18	0.16	>0.05	...
Area qualitative	9.14	0.004	0.12	10.9	0.02	0.14	0.12	>0.05	...
Area quantitative	13.5	$<10^{-4}$	0.17	0.16	>0.05	...	2.60	0.12	0.04

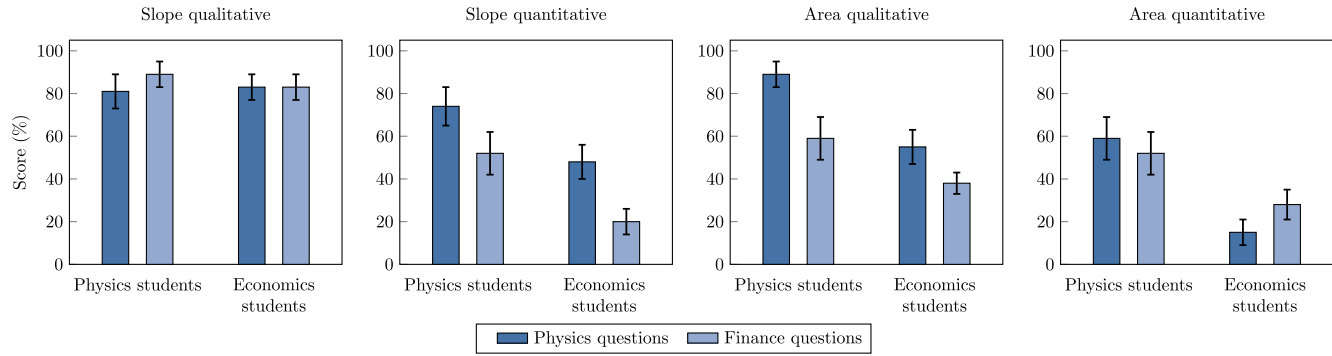


FIG. 4. Average scores of physics and economics students in the context of physics and finance on the qualitative and quantitative questions about graph slope and area under a graph. The error bars represent 1 SEM.

Appendix). To evaluate the impact of context on students' confidence ratings, a repeated-measure ANOVA was run with context as the within-subject factor and group as the between-subject factor for each pair of isomorphic questions. The results are shown in Table III in the Appendix. There were no significant effects for any combination of question type and concept, neither main effects on any factor nor interaction effects.

To investigate the accuracy of students' confidence judgments, the confidence ratings were considered for correct and incorrect answers, respectively. This dichotomous split of the dataset prevents the application of the repeated-measure analysis from above due to the lack of paired variables (there is only one confidence rating for either a correct or an incorrect answer). Hence, all questions on the slope concept and on the area concept were aggregated, respectively (Fig. 5). For the slope concept, the physics students were more confident when answering correctly than incorrectly [$t(106) = 2.62$, $p < 0.01$]. In contrast, the economics students' confidence was not significantly different between correct and incorrect answers [$t(145) = 1.37$, $p > 0.05$]. For the area concept, the results were similar. Physics students responded with higher confidence when

they answered correctly [$t(106) = 2.23$, $p < 0.05$], but economics students did not [$t(158) = -1.22$, $p > 0.05$].

C. Analysis of students' eye movements

1. Total viewing time

First, we calculated total viewing time, i.e., dwell time that students spent on all questions before giving the answer and before rating their confidence. Average total viewing time on all eight questions was 424 ± 90 s for the physics students and 458 ± 160 s for the economics students, with no significant difference between the groups. To compare students' total viewing time on qualitative and quantitative questions about graph slope and the area under a graph, a 2×2 ANOVA (type of question \times concept) was conducted, separately for both groups of students. For physics students, a significant main effect for concept was found [$F(1, 26) = 12.7$, $p < 0.001$, $\eta_p^2 = 0.33$], whereas the factor question type was not significant. Physics students spent more time viewing the area questions than viewing the slope questions [Fig. 6(a)]. The interaction effect was significant [$F(1, 26) = 10.9$, $p < 0.01$, $\eta_p^2 = 0.30$], indicating that the type of question

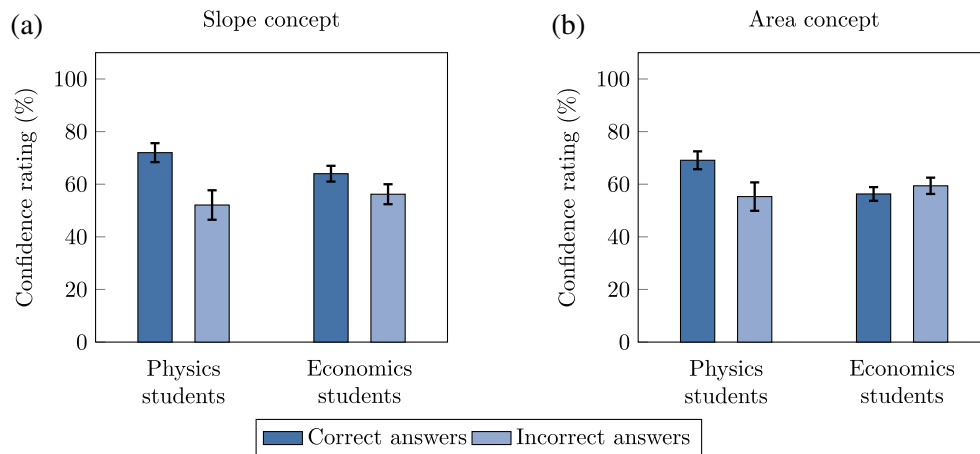


FIG. 5. Average confidence ratings of physics and economics students associated with correct and incorrect answers on (a) the slope questions and on (b) the area questions. The error bars represent 1 SEM.

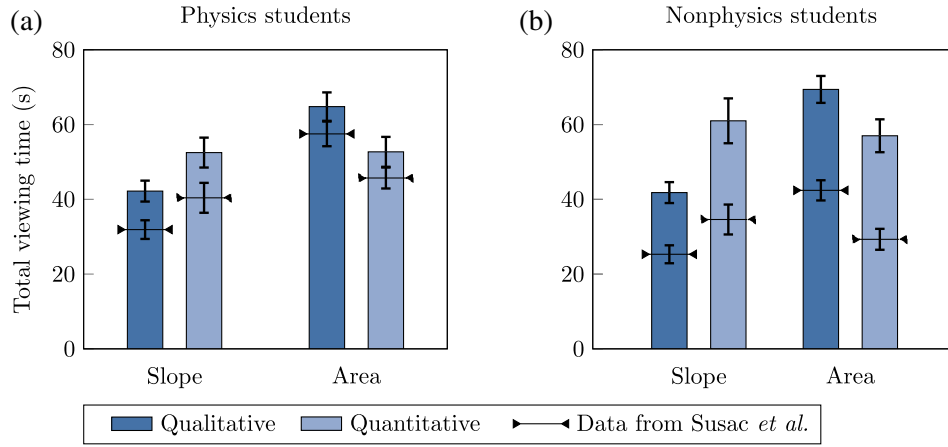


FIG. 6. Average total viewing time of (a) physics students and (b) nonphysics students on the qualitative and quantitative questions about graph slope and the area under a graph. The bar graphs represent data obtained in this study and refer to physics and economics students, whereas the triangles represent the data from Susac *et al.*, referring to physics and psychology students [7]. The error bars represent 1 SEM.

had the opposite effect on questions about graph slope than on questions about the area under a graph. Regarding the questions about graph slope, physics students paid more visual attention to the quantitative questions [$t(26) = 2.3$, $p < 0.05$], whereas they had longer viewing times for the qualitative questions [$t(26) = 2.1$, $p < 0.05$] on questions about the area under the graph.

For economics students, the results were analogous; we found a significant main effect of concept [$F(1, 39) = 12.9$, $p < 0.001$, $\eta_p^2 = 0.25$] and an interaction effect [$F(1, 39) = 20.9$, $p < 10^{-4}$, $\eta_p^2 = 0.35$] but no effect of question type. Economics students spent more time on questions about the area under a graph than they spent on questions about graph slope [Fig. 6(b)]. Similar to physics students, economics students had longer total viewing times for the quantitative questions about slopes [$t(39) = 3.62$, $p < 0.001$], whereas they had longer total viewing time for the qualitative questions about the area under a graph [$t(39) = 2.77$, $p < 0.01$].

To further explore students' total viewing times, we applied the same analysis that we applied to the scores and the confidence ratings. The results of a two-way mixed design ANOVA with the between-subject factor group and the within-subject factor context are shown in Table IV in the Appendix for each pair of isomorphic questions. The analysis revealed no effect of context or group in any combination of question type and concept. For qualitative questions on slopes, there was an interaction effect, indicating that physics students spent more time on the finance question and economics students spent more time on the physics question. Other interaction effects did not occur.

Finally, we compare our data to the total viewing time data from Susac *et al.* [7]. As Fig. 6 shows, the German physics and economics students spent more time viewing the items than the Croatian physics and psychology students did for all combinations of questions and concepts. The difference is

caused by viewing time on the question and the alternatives; see Sec. III C 2. However, the descriptive data in Fig. 6 also reveal a correlation between the viewing times and the concept or question type across both samples. In fact, all effects of question type and concept on students' total viewing time that Susac *et al.* reported have been reproduced in our study. As reported above (cf. Table IV), we found an interaction effect between study domain and context regarding total viewing time on the qualitative questions about graph slope. Besides that, no differences between the groups of students across the context were found. In contrast, Susac *et al.* reported two main effects of the factor group (psychology students spent less time on the area questions than physics students did) and two main effects on the factor context (students spent more time on the quantitative questions with a finance context than with a physics context).

2. Viewing time on different areas of interest

To further investigate the eye-tracking data, the viewing times were analyzed with respect to different areas of interest, as defined in Sec. II E. First, we obtained students' viewing times on the AOIs *Question*, *Graph*, and *Multiple choice* and compared them between the groups of students (Fig. 7). Three Bonferroni adjusted t -tests showed no statistical difference between the viewing time of physics and economics students at the AOIs *Question* [$t(65) = -0.26$, $p > 0.05$], *Graph* [$t(65) = -0.74$, $p > 0.05$], and *Multiple choice* [$t(65) = -0.54$, $p > 0.05$]. In contrast to our result, Susac *et al.* reported a significant difference between the viewing time of physics and nonphysics students at the AOI *Graph*; see Fig. 7. While the German and Croatian physics students spent the same time on the AOI *Graph*, there is a huge difference between psychology and economics students. From Fig. 7 it also seems that Croatian students spent less time than German students on the *Question* and *Multiple choice* AOIs.

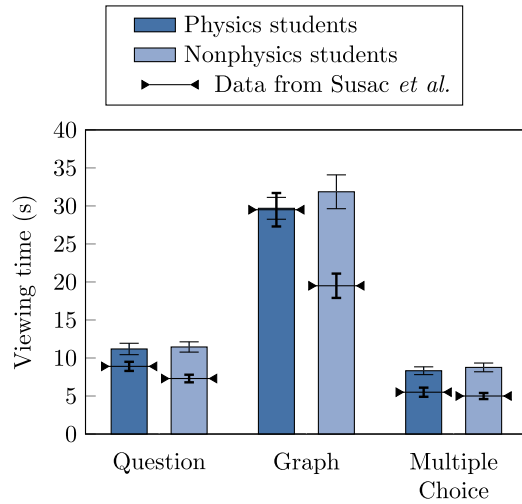


FIG. 7. Average viewing time of physics and nonphysics students on the AOIs *Question*, *Graph*, and *Multiple choice*. The bar graphs represent data obtained in this study and refer to physics and economics students, whereas the triangles represent the data from Susac *et al.*, referring to physics and psychology students. The error bars represent 1 SEM.

Second, the total viewing time on the AOI *Axis labels* (adding the viewing times on the *x*-axis and *y*-axis labels) was determined for each item. We performed a two-way mixed design ANOVA with the between-subject factor group and the within-subject factor context on total viewing time on the AOI *Axis labels* and found a significant interaction [$F(1, 65) = 6.9$, $p = 0.01$, $\eta_p^2 = 0.10$]. For questions about the physics context, physics students looked at the axis labels for 14 ± 9 s in total while the total viewing time tended to be longer for economics students (19 ± 11 s). The opposite was the case for questions about the finance context: economics students had slightly shorter total viewing times (15 ± 7 s) than physics students had (18 ± 9 s) [33]. Considering the AOI *Axis labels*, Susac *et al.* also found a statistically significant interaction between the factors group and context. Croatian physics students spent more time on the finance labels whereas the psychology students spent more time on the physics labels.

Third, the viewing times on the *Axis tick labels* were analyzed by a 2×2 ANOVA (type of question \times concept), separately for the physics and the economics students. For physics students, there was a significant main effect of the factor question type [$F(1, 26) = 24.4$, $p < 10^{-4}$, $\eta_p^2 = 0.48$]. Physics students spent more time on the axis tick labels of quantitative questions; see Fig. 13 in the Appendix. For the economics students, both main effects were significant, the factor concept [$F(1, 39) = 40.5$, $p < 10^{-4}$, $\eta_p^2 = 0.51$] and the factor question type [$F(1, 39) = 13.4$, $p = 0.001$, $\eta_p^2 = 0.23$]. Economics students also had more fixations on the axis tick labels of quantitative questions and on the axis tick labels of questions about the area under the graph.

3. Distribution of visual attention and saccadic eye movements

Last, we analyzed the eye-tracking data on the item level. We determined the distribution of fixations and the distribution of absolute saccade directions for physics and economics students with respect to the *Graph* (without the axes labels). The analysis was performed for each item, and the results of one item per question type and concept is reported in this article. The analysis of the remaining items is reported in the Appendix. Figures 8(a), 9(a), 10(a), and 11(a) present the differences in visual attention between the groups. If the physics students had more fixations on a specific AOI G_i than the economics students, the AOI is colored green. Otherwise, it is colored red. Uncolored AOIs reflect areas with no difference in visual attention between physics and economics students, occurring either if there was no visual attention at all or both groups of students had the same number of fixations.

Figures 8(b), 9(b), 10(b), and 11(b) show the distribution of absolute saccade directions in physics and economics students. The distribution of saccadic directions was generated by counting the number of saccadic directions

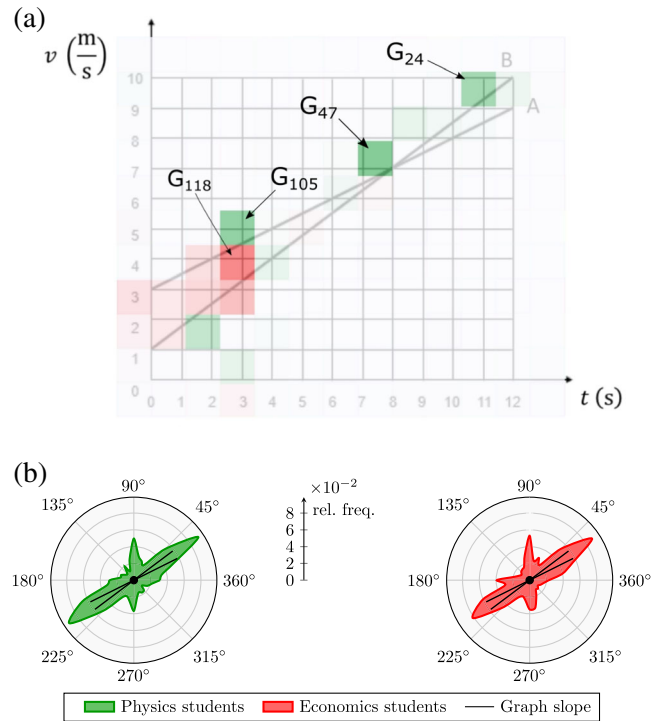


FIG. 8. Analysis of visual attention and saccadic eye movements for the qualitative question on graph slope in the context of physics. Students had to compare acceleration of two objects within the first 3 s of the motion process. (a) The green (red) color indicates areas that were observed significantly longer by physics (economics) students. (b) Polar distributions of saccadic angles by group condition (physics students vs economics students), referring to the AOI *Graph*. The solid black lines refer to the gradient angles of the line graphs.

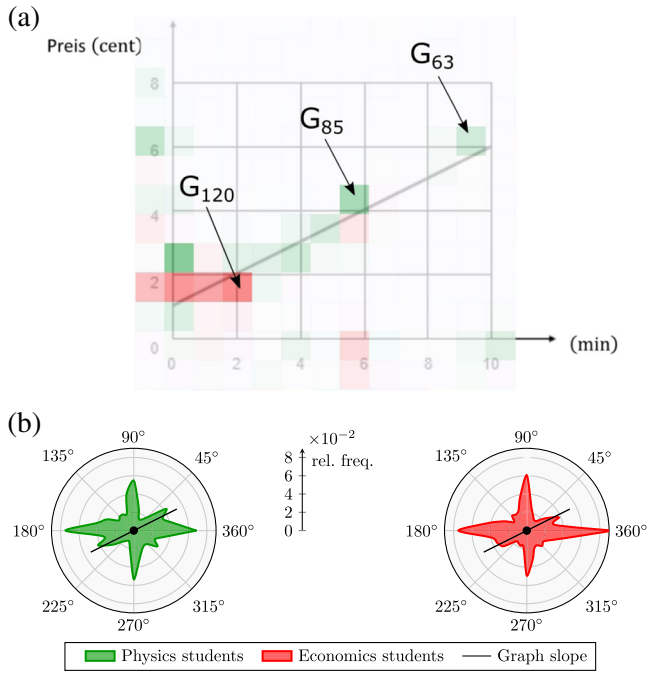


FIG. 9. Analysis of visual attention and saccadic eye movements for the quantitative question on graph slope in the context of finance. Students had to determine the cost per minute from the graph. For more details, please refer to the caption of Fig. 8.

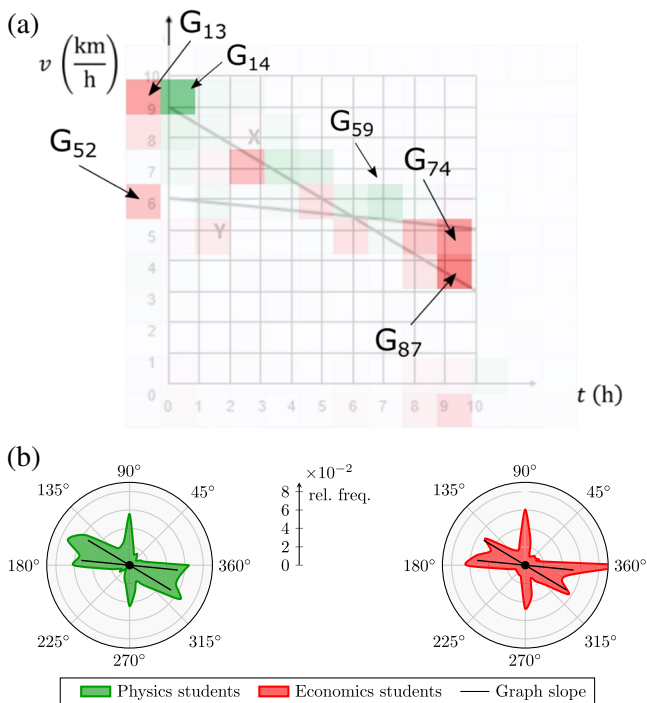


FIG. 10. Analysis of visual attention and saccadic eye movements for the qualitative question on area under a graph in the context of physics. Students had to compare the distances covered by two objects after $t = 9$ h. For more details, please refer to the caption of Fig. 8.

within each bin of the full circle and by normalizing to the total number of saccades. Furthermore, each plot shows the gradient angles of the line graphs in unit length, allowing a convenient comparison of the students' saccadic eye movements to the slope of the line graphs. The distribution of absolute saccade directions tells us whether students followed the graph or made transitions between numbers on the vertical or horizontal axes and the graph.

The qualitative question about graph slope requires a comparison between the acceleration of two objects within the first 3 s of the motion process. Economics students show a tendency to evaluate the lower left part of the diagram with more attention than physics students did [Fig. 8(a)], whereas physics students also focus on regions of the graph that are not addressed by the question (G_{24} , G_{47} , G_{105}). In particular, the AOI G_{118} covers the area between both line graphs at $t = 3$ s and receives more attention from economics students than from physics students. The polar distribution of saccade directions shows peaks at angles that correspond to the gradient angle of the graphs. In numbers, more than 40% of eye movements were performed in the same direction as the line graphs (within a 5° tolerance band). Thus, the saccadic direction corresponding to the line graph gradient is the most dominant angle in the distribution of physics and economics students, indicating that both groups of students followed the line graphs with their eyes.

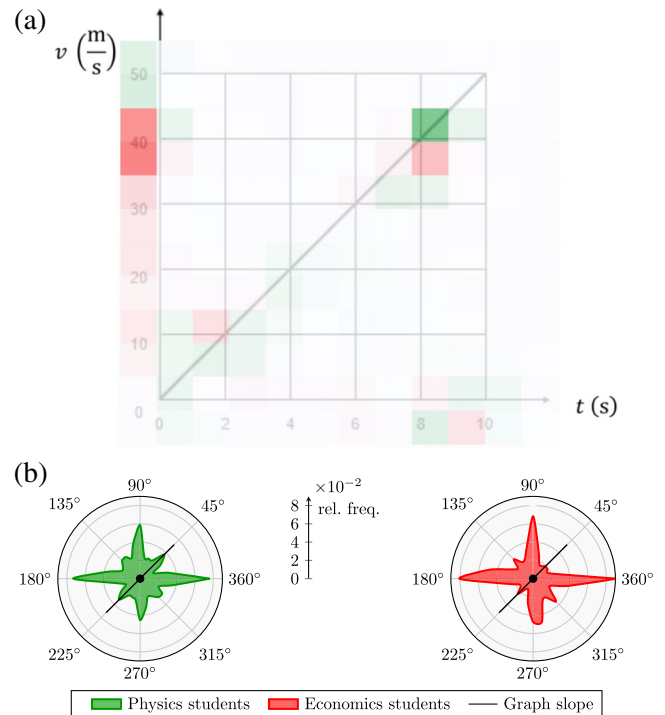


FIG. 11. Analysis of visual attention and saccadic eye movements for the quantitative question on area under a graph in the context of physics. Students had to calculate the distance traveled by a car during the first 8 s from the graph. For more details, please refer to the caption of Fig. 8.

For the quantitative question about graph slope in the finance context, students had to calculate the cost per minute of a phone call. As can be seen in Fig. 9, economics students focused relatively more on the point (2,2) (AOI G_{120}) than the physics students did. On the contrary, the physics students allocated relatively more visual attention to different intersection points, e.g., (6,4) (AOI G_{85}) and (10,6) (AOI G_{63}), and they fixated the graph relatively more often than economics students did. Figure 9 also reveals that economics students fixated the point (6,4) and the corresponding axis tick labels but less often than the physics students on average. The saccadic direction distribution is similar between the groups, peaking in horizontal and vertical directions. We found that physics students performed more saccadic eye movement along the direction of the graph slope (26.6° and 206.6°) than economics students did [$t(65) = 2.35$, $p = 0.02$].

The qualitative question about the area under a graph asked students to compare the distances covered by two objects after $t = 9$ h in a $v(t)$ diagram. The most pronounced difference in visual attention between physics and economics students can be found in the AOIs G_{74} and G_{87} , covering the graph line at $t = 9$ h; see Fig. 10. Economics students spent much more time on these areas than did physics students. Furthermore, economics students spent longer time focusing on the intersection of the line graphs with the y axis (G_{13} and G_{52}). Physics students focused mostly on the graph “X”, and predominantly in regions where the graph “X” is above graph “Y” (G_{14} , G_{59}). The saccadic direction distributions show peaks in vertical direction and in the direction of the graph slope. While the distributions are very similar between both groups of students, economics students tended to perform more horizontal eye movements than physics students did [$t(65) = 1.70$, $p = 0.10$].

Finally, the quantitative question about the area under a graph required students to calculate the distance traveled by a car during the first 8 s, given a $v(t)$ diagram. Economics students spent more time looking at the values on the y axis than physics students did, see Fig. 11, whereas physics students distributed their attention more along the graph line, focusing on the point (8,40). The saccadic direction distribution shows peaks in horizontal and vertical directions. Furthermore, physics students performed more saccadic eye movement along the direction of the graph slope (45° and 225°) than economics students [$t(65) = 2.20$, $p = 0.03$].

To compare the distribution of visual attention among isomorphic pairs of items, please refer to Fig. 14 in the Appendix, which reinforces the results described above.

IV. DISCUSSION

A. Students’ scores and confidence ratings

1. Physics students achieved better scores than economics students

Physics students performed significantly better than economics students overall. Apart from qualitative questions

about graph slope, physics students had much higher scores in all combinations of question type and concept. Thus, we replicated the findings from Susac *et al.* [7] with another nonphysics comparison group, viz., economics students, and with a physics sample that is substantially different from the original sample (first-year physics majors vs fourth-year prospective physics teachers).

2. Physics students assess their understanding more accurately than economics students

Physics students provided correct answers with higher confidence ratings in comparison to when they gave incorrect answers. Thus, the students’ ability and their confidence were highly intercorrelated. In contrast, the economics students often did not judge their cognitive skills accurately since there is no significant difference between the mean confidence ratings assigned to items which the students answered correctly and the mean confidence ratings assigned to items answered incorrectly. However, the physics students’ average confidence rating for incorrect answers is still about 50%, revealing potential to improve their metacognitive skills. Several investigations found that the accuracy of confidence judgments varies among different ability levels; viz., low performing students do not reliably distinguish between correct and incorrect responses [9,11,34,35]. This phenomenon is well studied in educational psychology and is called the Dunning-Kruger effect [13]. In this particular case of items, incorrect alternatives were constructed to address common misconceptions and rather simple calculation errors (e.g., determining the area of a rectangle instead of a triangle in the case of the area questions). Therefore, the participants were able to select an incorrect alternative that reflected their flawed thinking process, resulting in a high response confidence [14]. It can be assumed that the mathematical procedures for solving the questions are more present for the physics than for the economics students [7], enabling physics students to reflect on their answers more deeply whereas the metacognitive insight is reduced for economics students.

3. Graph slope is an easier concept than the area under a graph

Similar to Susac *et al.* [7], all students solved the questions about graph slope better than the questions about the area under a graph. In particular, the qualitative questions about graph slope were solved correctly by around 80% of the students across both contexts, indicating that this idea is intuitive for physics and economics students (e.g., in the context of consumer surplus or producer surplus). The economics students also had no difficulties with the qualitative question of graph slope in the context of physics. This indicates that economics students were able to recognize acceleration as the slope in the v vs t graph in the same manner as the growth rate of prices as the slope in the price vs time graph.

4. Quantitative questions are more difficult than qualitative questions for first-year physics students

In contrast to the Croatian study, we found a main effect of question type (quantitative vs qualitative). The German physics students solved the qualitative questions better than the quantitative ones; i.e., the first-year students have more trouble with the quantitative problems compared to the qualitative problems whereas the fourth-year students investigated by Susac *et al.* solved both types of problems equally well. Similar results of this kind have been reported before [5] for first-year students, so it is likely that the difference between the results is due to the differences between the two physics samples. As Susac *et al.* point out, “studying physics significantly improves their [the students’] ability to solve quantitative problems on graphs” (Ref. [7], p. 10), and additional follow-up tests would shed more light on the development of this skill.

5. Transfer to familiar and unfamiliar domains

Physics students solved the physics questions better than the finance questions, which is not surprising given that they chose physics as their study domain. Moreover, they also solved the finance questions very well and even better than the economics students. Physics students probably never encountered these types of questions before; thus our results indicate that physics students were able to successfully transfer the mathematical strategies they developed in physics or mathematics at school for solving tasks in a different context.

For the economics students, we have no reliable data about their familiarity in the domains of physics or finance questions apart from their choice of the study domain (economics) and the small fraction of students that selected physics in school. Further studies should take this variable into account to obtain a better discrimination in terms of the students’ familiarity with each context. Counterintuitively, the economics students only performed better in the finance context compared to the physics context in one out of four pairs of questions (quantitative question about the area under a graph). This again raises the question about the familiarity of economics students in the finance context, and further studies are required to investigate the role of context familiarization as a prerequisite of performance.

B. Students’ eye movements

1. Total viewing time indicates item difficulty

Our findings broadly confirm the results obtained by Susac *et al.* that both groups of students spent more time on questions about the area under the graph than on questions about graph slope. This substantiates the evidence of the area concept being more difficult for the students compared to the slope concept, since longer viewing times are associated with higher cognitive effort [36].

Furthermore, we confirmed that solving quantitative questions about the slope is more time consuming than

solving the qualitative question. Susac *et al.* explain the different viewing times by the fact that students need to extract more information from the graph when calculations have to be performed. Indeed, we found that the longer viewing time on quantitative questions about graph slope can mostly be credited to longer viewing times on the axis tick labels, supporting the idea that information extraction and processing contributes to the difference, and extending previous results.

Judging the area under the graph qualitatively required the longest time across all items. It took students even longer than calculating the area under the graph quantitatively, which is an opposite trend of the viewing times comparing the slope qualitative and quantitative questions. The result is again well in line with the literature and can be explained by the fact that, first, the area under a graph cannot be estimated as quickly as the graph slope and, second, these types of questions were likely to be new for both groups of students, so maybe they needed more time to evaluate what to look for and where. For qualitative questions, both groups spent more time on the axis tick labels for the area question compared to the slope question. We assume that the students tried to get more information about the units of the quantities that were presented on the labels and how to combine them to fit the answers. Since more work has to be done to confirm this claim (e.g., using think-aloud interviews), the result at least shows that the demands of the qualitative questions on area were unfamiliar to the students.

2. Unfamiliar axis labels receive more attention

Our data reproduced the finding of Susac *et al.* that physics students spent more time on the finance axis labels and nonphysics students spent more time on the physics axis labels. Thus, both groups of students needed more time to extract information from the axes in graphs with the context that was unfamiliar to them. This result shows that both groups of students made an effort to extract relevant information contained in the graph axes and that the axis label is important to solve the questions. This result also serves as a secondary evaluation of the experiment (“manipulation check” [37]), showing that each student group was at least more familiar with the terminology used in their own domain.

3. Total viewing time is not correlated with the students’ performance

The comparison between physics and economics students revealed no differences between the groups concerning total viewing times, whereas Susac *et al.* report differences between physics and psychology students [7]. The results further showed that the German physics students spent more time on the items than the Croatian physics students did, and that the difference arises from a longer viewing time on the question statement (*Question*) and the alternatives (*Multiple choice*). More importantly, both samples of physics students as well as the economics students spent about the same amount of time on the graph (*Graph*). The difference in total

scores between physics $[(69 \pm 26)\%]$ and economics students $[(46 \pm 23)\%]$ is not reflected by differences in total viewing time on the diagrams; thus, total viewing time alone does not explain the difference in the performance outcomes between the groups.

4. Economics students spend more time on conceptually irrelevant areas than physics students

The pattern analysis revealed several manifestations of misconceptions and learning difficulties for economics students on the item level. When required to calculate the slope (e.g., velocity from a position-time graph) at a given point, students oftentimes divide the y coordinate (say, the position) by the x coordinate (say, time) instead of considering appropriate intervals (*point-interval confusion*). For quantitative questions about graph slope, the economics students spent relatively more time on a single point in the diagram $[(x, y) = (2, 2)]$ than the physics students did, who in turn considered many different points. Indeed, most incorrect answers included a division of two values y/x rather than the division of intervals $\Delta y/\Delta x$. Prior literature reported about the interval-point confusion as a common misconception in the context of graph understanding [4,38], and the eye-tracking data provide evidence for this mistake. Susac *et al.* also reported that using only points instead of the intervals was the most common incorrect strategy used by the students [7].

For qualitative questions about the area under the graph, the economics students spent relatively more time on the axis tick labels even though they were not relevant to estimate the area under the curves. Beichner reported that students inappropriately use axis values when area calculations are required and our data support this claim [1]. We also found that economics students relative to physics students focused more often on areas that were highlighted by the question statement (e.g., “ $t = 9$ h” for the qualitative question about area and “ $t = 3$ s” for the qualitative question about slope); see Figs. 8 and 10. Each question was phrased to address a specific point in the graph (e.g., to calculate the slope at the instant $t = 3$ s). However, focusing only on this instant furthermore indicates the point-interval confusion.

We found that economics students also showed a tendency to compare the graph height rather than comparing the area under the curve (Figs. 10 and 11). Students do not recognize the meaning of areas under kinematics graph which has also been reported before [1], and the eye-tracking data prove its existence.

5. Physics students trace the line graphs with their eyes more often than economics students

For qualitative questions, the saccadic direction distribution showed that both groups of students performed saccadic eye movements corresponding to the gradient angle of the line graph. For the qualitative questions about graph slope, most eye movements followed the graph. We interpret this result as an indication of correct cognitive processing of the

slope concept by the students in a qualitative manner. This suggests that the slope concept is an intuitive idea and that both groups of students activate similar cognitive resources to solve the problem. For the quantitative questions, horizontal and vertical eye movements were dominant, indicating interactions between points on the graph and the axis tick labels, and comparing values on the horizontal (x) axis and vertical (y) axis. Therefore, the distribution of saccadic directions is a discriminating feature of question type for graph slope (quantitative vs qualitative).

Comparing physics and economics students, the analysis of saccadic directions revealed a significant difference between the groups for the quantitative questions. Physics students followed the line graphs with their eyes more often than economics students did, and this difference becomes more obvious when fixations on the graph axis and on axis tick labels are excluded from the analysis. This supports the finding from above that physics students distributed their attention over a broader area and scanned the line graphs more often to solve the problems more successfully.

V. CONCLUSION

The aim of this study was to compare physics and nonphysics students regarding their understanding of graph slope and the area under the graph in the contexts of physics and finance. In particular, we replicated the investigation performed by Susac *et al.* [7] with a different nonphysics sample (economics students). The thorough eye-tracking analysis sheds more light on differences between physics and nonphysics students, reflecting correct and incorrect strategies that have been identified in the literature.

Overall, the physics students outperformed the economics students, especially in the context of finance questions. Even more, economics students overestimated their ability and we hypothesize that this effect can be attributed to the familiarity with the finance context. However, further investigations are required to reliably assess context familiarity of different student groups. Since students might be more able to develop an appropriate understanding of graphs if they reflect correctly on their performance [9], the evaluation of and confrontation with confidence judgments could furthermore help them to identify knowledge gaps. This confirming observation supports the demand of Susac *et al.* for a substantial mathematical education in high school and college level education in various fields, since it appears to be essential not only to transfer the skills to an unfamiliar domain but also to understand and correctly solve problems in the familiar domain. In light of these results, we suggest that future research address the development of instruction with respect to different contexts in order to promote students' abilities to transfer their strategies to different domains.

There were no differences concerning time spent on the diagrams between physics and economics students, proving this measure to be inadequate for discriminating between the different performance scores of the groups. Instead, more advanced analysis methods revealed differences in the visual attention between both groups of students, reflecting

typical misconceptions such as the point-interval confusion and confusion between absolute value and slope or area. The identification of conceptually relevant and irrelevant areas paves the way for guided-feedback systems and machine-learning algorithms in the context of graph understanding [39] and can be used for discussions in the classroom. In particular, the typical mistakes should be addressed and sample gaze paths from successful physics students can be used to demonstrate and discuss expert information extraction from diagrams.

Furthermore, our results broadly confirm previous findings on student understanding of graphs; i.e., graph slope is an easier concept than the area under the graph for physics and nonphysics students. Area questions required more time and were therefore cognitively more demanding, indicating that more emphasis should be put on the qualitative and quantitative evaluation of the area concept.

Extending previous results, we found that qualitative and quantitative questions are cognitively processed differently. Students spent more time on the axis tick labels and predominantly perform saccades in horizontal and vertical directions, i.e., along the axes, when they encounter quantitative questions. In contrast, students predominantly performed saccades corresponding to the gradient angle of line graphs to extract qualitative information.

ACKNOWLEDGMENTS

The authors thank Professor Georg von Freyman (TUK) for welcoming this research in his course and Sebastian Becker for helpful discussions. This work is funded by the Federal Ministry of Education and Research (BMBF; project: VorleXung; supportcode: 16DHL1001). We would also like to thank the Rhine-Main Universities (RMU) Initiative Fund for supporting this study.

APPENDIX: ADDITIONAL DATA

The students' confidence ratings are depicted in Fig. 12, and the belonging two-way ANOVA results are summarized in Table III. Additional eye-tracking data referring to the AOI Axis tick labels are provided in Fig. 13. Finally, the results of the saccadic direction analysis for items not discussed in Sect. III C 3 are provided in Fig. 14. This figure also contains the students' spatial distribution of visual attention.

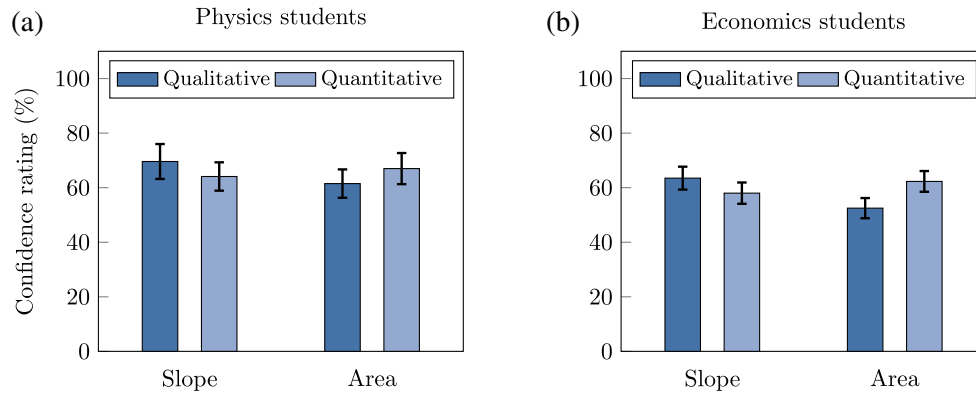


FIG. 12. Average confidence ratings of physics students (a) and nonphysics students (b) on the qualitative and quantitative questions about graph slope and area under a graph. The error bars represent 1 SEM.

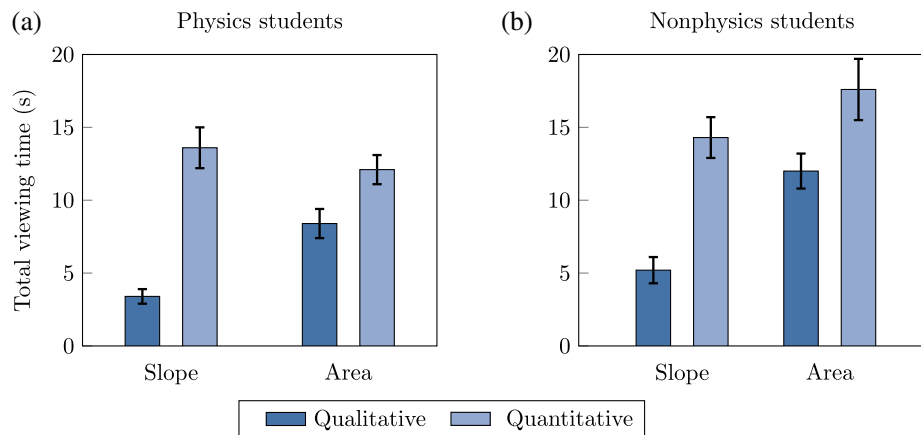


FIG. 13. Average total viewing time of (a) physics students and (b) nonphysics students on the AOI Axis tick labels for qualitative and quantitative questions about graph slope and area under a graph. The error bars represent 1 SEM.

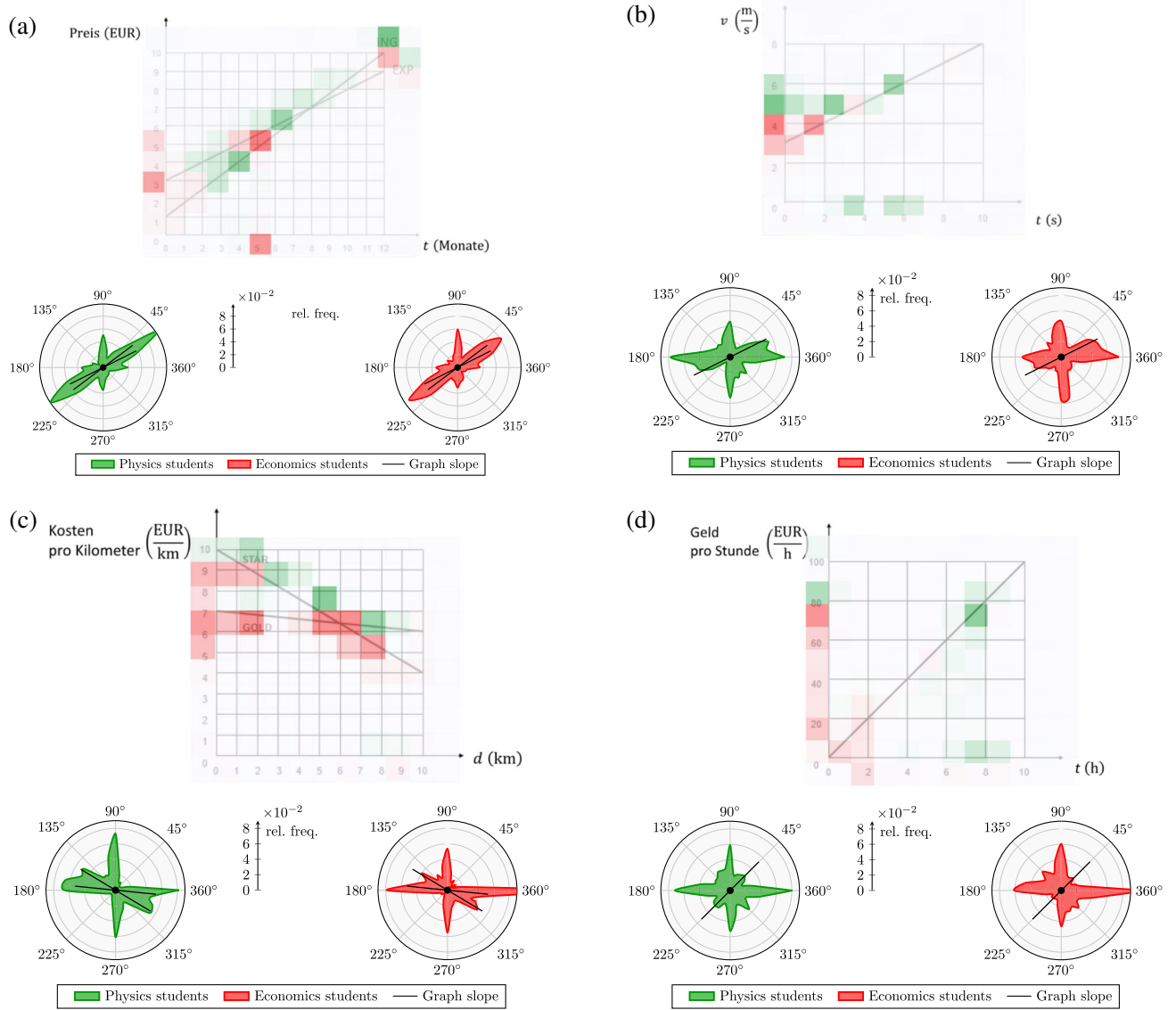


FIG. 14. Analysis of visual attention and saccadic directions for (a) the qualitative question on graph slope in the context of finance, for (b) the quantitative question on graph slope in the context of physics, for (c) the qualitative question on area under a curve in the context of finance, and for (d) the quantitative question on area under a curve in the context of finance. The green (red) color indicates areas that were observed significantly longer by physics (economics) students. In (a), students had to compare price rates of two objects within the first five months of a time period. In (b), students had to determine the acceleration from the graph. In (c), students had to compare the total costs for a taxi ride (8 km) between two companies. In (d), students had to calculate from the diagram the total salary of a worker after working 8 hours.

TABLE III. Results of the two-way ANOVAs conducted on students' confidence ratings with the context (physics vs finance) as a within-subject factor and with the group (physics students vs economics students) as a between-subject factor.

	Group			Context			Interaction		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Slope qualitative	0.70	>0.05	...	0.14	>0.05	...	0.97	>0.05	...
Slope quantitative	0.90	>0.05	...	0.04	>0.05	...	0.50	>0.05	...
Area qualitative	2.11	>0.05	...	0.06	>0.05	...	0.25	>0.05	...
Area quantitative	0.51	>0.05	...	0.02	>0.05	...	2.00	>0.05	...

TABLE IV. Results of the two-way ANOVAs conducted on total viewing time with the context (physics vs finance) as a within-subject factor and with the group (physics students vs economics students) as a between-subject factor.

	Group			Context			Interaction		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Slope qualitative	0.01	>0.05	...	0.65	>0.05	...	11.1	0.001	0.15
Slope quantitative	1.12	>0.05	...	1.63	>0.05	...	0.38	>0.05	...
Area qualitative	0.74	>0.05	...	0.41	>0.05	...	2.00	>0.05	...
Area quantitative	0.46	>0.05	...	2.5	>0.05	...	0.21	>0.05	...

- [1] R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [2] M. C. Linn, J. W. Layman, and R. Nachmias, Cognitive consequences of microcomputer-based laboratories: Graphing skills development, *Contemp. Educ. Psychol.* **12**, 244 (1987).
- [3] T. Wemyss and P. Van Kampen, Categorization of first-year university students' interpretations of numerical linear distance-time graphs, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010107 (2013).
- [4] G. Leinhardt, O. Zaslavsky, and M. K. Stein, Functions, graphs, and graphing: Tasks, learning, and teaching, *Rev. Educ. Res.* **60**, 1 (1990).
- [5] M. Planinic, L. Ivanjek, A. Susac, and Z. Milin-Sipus, Comparison of university students' understanding of graphs in different contexts, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020103 (2013).
- [6] J. Tuminaro, A cognitive framework for analyzing and describing introductory students' use and understanding of mathematics in physics, Ph.D. thesis, University of Maryland, 2004.
- [7] 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).
- [8] S. Hasan, D. Bagayoko, and E. L. Kelley, Misconceptions and the certainty of response index (CRI), *Phys. Educ.* **34**, 294 (1999).
- [9] B. A. Lindsey and M. L. Nagel, Do students know what they know? Exploring the accuracy of students' self-assessments, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020103 (2015).
- [10] 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).
- [11] S. Brückner and O. Zlatkin-Troitschanskaia, in *Assessment of Learning Outcomes in Higher Education. Methodology of Educational Measurement and Assessment*, edited by O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, and C. Kuhn (Springer Publishing, Cham, 2018), pp. 101–123, https://doi.org/10.1007/978-3-319-74338-7_6.
- [12] J. J. Shaughnessy, Confidence-judgment accuracy as a predictor of test performance, *J. Res. Personality* **13**, 505 (1979).
- [13] J. Kruger and D. Dunning, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments, *J. Pers. Soc. Psychol.* **77**, 1121 (1999).
- [14] P. Klein, A. Müller, and J. Kuhn, Assessment of representational competence in kinematics, *Phys. Rev. Phys. Educ. Res.* **13**, 010132 (2017).
- [15] 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).
- [16] M. Kozhevnikov, M. A. Motes, and M. Hegarty, Spatial visualization in physics problem solving, *Cogn. Sci.* **31**, 549 (2007).
- [17] 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).
- [18] 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).
- [19] M. Kekule, Students' different approaches to solving problems from kinematics in respect of good and poor performance, in *Proceedings of the International Conference on Contemporary Issues in Education, Dubai* (2015), pp. 126–34.
- [20] M. Kekule, Students' approaches when dealing with kinematics graphs explored by eye-tracking research method, in *Proceedings of the Frontiers in Mathematics and Science Education Research Conference (FISER)* (2014), pp. 108–117, <http://scimath.net/fiser2014/presentations/Martina%20Kekule.pdf>.
- [21] D. T. Lykken, Statistical significance in psychological research, *Psychol. Bull.* **70**, 151 (1968).
- [22] S. Schmidt, Shall we really do it again? The powerful concept of replication is neglected in the social sciences, *Rev. Gen. Psychol.* **13**, 90 (2009).

- [23] K. Cummings, J. Marx, R. Thornton, and D. Kuhl, Evaluating innovation in studio physics, *Am. J. Phys.* **67**, S38 (1999).
- [24] J. W. Belcher, Improving student understanding with TEAL, MIT Faculty Newsletter **16**, 8 (2003).
- [25] A. P. Fagen, C. H. Crouch, and E. Mazur, Peer Instruction: Results from a range of classrooms, *Phys. Teach.* **40**, 206 (2002).
- [26] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- [27] M. C. Makel and J. A. Plucker, Facts are more important than novelty: Replication in the education sciences, *Educ. Res.* **43**, 304 (2014).
- [28] See Supplemental Material in 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), <https://journals.aps.org/prper/supplemental/10.1103/PhysRevPhysEducRes.14.020109>.
- [29] More specifications can be found on the product website <https://www.tobiipro.com>.
- [30] D. D. Salvucci and J. H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Association for Computing Machinery, New York, 2000), pp. 71–78.
- [31] The accuracy of the system was about 0.4° visual angle, which corresponds to $a = 0.4$ cm on the screen given the average distance from eye to screen of 60 cm. For the analysis, each rectangle covered 50×50 pixels. With a 22" screen and a resolution of 1920×1080 pixels, each rectangle covered an area of $1.25 \text{ cm} \times 1.25 \text{ cm} = 1.56 \text{ cm}^2 \gg a^2 = 0.16 \text{ cm}^2$.
- [32] 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).
- [33] Note that the differences in each subgroup were not significant even though the interaction effect was significant.
- [34] P. Bell and D. Volckmann, Knowledge surveys in general chemistry: Confidence, overconfidence, and performance, *J. Chem. Educ.* **88**, 1469 (2011).
- [35] M. D. Sharma and J. Bewes, Self-monitoring: Confidence, academic achievement and gender differences in physics, *J. Learn. Des.* **4**, 1 (2011).
- [36] 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).
- [37] M. Allen, *The SAGE Encyclopedia of Communication Research Methods* (SAGE Publications, Thousand Oaks, CA, 2017), Vols. 1–4.
- [38] L. C. McDermott, M. L. Rosenquist, and E. H. van Zee, Student difficulties in connecting graphs and physics: Examples from kinematics, *Am. J. Phys.* **55**, 503 (1987).
- [39] N. Rebello, M. Nguyen, Y. Wang, T. Zu, J. Hutson, and L. Loschky, Machine learning predicts responses to conceptual questions using eye movements, in *Proceedings of the Physics Education Research Conference 2018, Washington, DC* (2018), <http://dx.doi.org/10.1119/perc.2018.pr.Rebello>.