9th grade students' understanding and strategies when solving $x(t)$ problems in 1D kinematics and $y(x)$ problems in mathematics

Stijn Ceuppens[*] and Laurens Bollen

KU Leuven, Department of Physics and Astronomy, Celestijnenlaan 200c box 2406, 3001 Leuven, Belgium

Johan Deprez

KU Leuven, Department of Mathematics, Celestijnenlaan 200b box 2400, 3001 Leuven, Belgium

Wim Dehaene

KU Leuven, ESAT—MICAS, Microelectronics and Sensors,
Kasteelpark Arenberg 10 box 2443, 3001 Leuven, Belgium

Mieke De Cock[†]

KU Leuven, Department of Physics and Astronomy, Celestijnenlaan 200c box 2406, 3001 Leuven, Belgium

We design, validate, and administer a 24-item test to study student understanding of linear functions in 1D kinematics $[x(t)]$ and mathematics $[y(x)]$ in the 9th grade. The items assess identification and comparison of initial position and velocity in 1D kinematics and of the y intercept and slope in mathematics using a graph or an algebraic formula. Results show that students' performance on most mathematics items is significantly better than on their isomorphic kinematic counterparts, but also that most of the easiest as well as the most difficult items are kinematics items. Students achieve the highest accuracies on graphical questions in which they must compare two positive slopes, and they achieve the lowest accuracies on questions in which they must determine or compare a negative slope. We find that students have more difficulties with the y intercept in mathematics and with the slope in kinematics. Furthermore, questions in symbolic representation result in far lower accuracies compared to questions in graphical representation, particularly when the y intercept or the slope has to be determined instead of compared. We also analyze the results qualitatively by categorizing the students' strategies and errors. We find frequent confusion between the x intercept and the y intercept in mathematics, but far less in kinematics. Negative velocities in kinematics are by far the largest pitfall, whereas negative slope in mathematics is rarely an issue. The results also show a significant frequency of interval or point confusions in kinematics but very little in mathematics. We reaffirm the occurrence of the interval or point confusion in questions with graphs and discuss three different cases of interval or point confusions in questions with algebraic expressions: numerical, algebraic, and unit based. Our results indicate a weak link between kinematics and mathematics and we suggest that closer integration between these two contexts during education could benefit student understanding of linear functions and linear phenomena in kinematics.

I. INTRODUCTION

Linear functions are an essential part of any science curriculum but recent studies show that students' understanding thereof is subpar and that the context is an important factor [1–8]. To gain more insight into students' performance and into their strategies to solve linear function problems during the early stages of learning about them, we investigate how 9th grade students in Flanders (Belgium) solve $x(t)$ problems in 1D kinematics and in as-isomorphic-as-possible $y(x)$ problems in mathematics. Furthermore, we are interested in the effect of the representation and the task on students' strategies and accuracy. The target group is particularly interesting since these students have only just received education on linear relations in mathematics and physics. However, the mathematics and the physics curricula are not explicitly aligned and uniform linear motion and linear functions are taught

[*]stijn.ceuppens@kuleuven.be
[†]mieke.decock@kuleuven.be

without much reference to each other. As such, it is unlikely that students have constructed a deep understanding of linear relations in different contexts and in different representations. This mismatch opens the door for a deeper look into how students link their understanding of linear functions between physics and mathematics, which are the two contexts in the curriculum likely to benefit the most from increased synergy between them. To do so, we develop and validate a test, and perform both quantitative and qualitative analysis of students' answers to gain as much insight as possible.

We first provide background information from literature about relevant factors in Sec. II. Next, we explain our research design and the test development in Sec. III, followed by a description of the test administration in Sec. IV. The quantitative and qualitative results are described in Secs. V and VI, respectively. These results are then jointly discussed in Sec. VII. Finally, implications for teaching and future research are discussed in Sec. VIII and a conclusion is provided in Sec. IX.

II. LITERATURE

Linear functions are of paramount importance to describe, understand, and approximate many phenomena in physics. Students encounter them throughout their education and they are often the first step in learning about abstract relationships between dependent and independent variables. In most curricula, students start learning about linear functions in lower secondary education. At that time, they are likely to already have some deep-rooted beliefs and tend to operate in different belief systems: the real world, the physics world, and the mathematics world, which are not always well linked [9]. Multiple studies on this topic show that conceptual understanding of linearity is indeed not easily obtained and many difficulties have been reported across various contexts and across various representational formats such as graphs and formulas [1–8,10–22]. The target groups in these studies range from secondary to higher education but the results are very similar, which is suggestive of some obstinacy for these difficulties throughout education. The slope of graphs of linear functions in physics has been extensively studied and the most common difficulties were grouped by Leinhardt *et al.* [13] into three distinct categories: slope or height confusion in which the height of the curve is mistaken for its slope; interval or point confusion in which a single point is considered when an interval is more appropriate, i.e., taking the ratio of coordinates instead of the ratio of intervals to determine the slope; or iconic interpretations in which the graph is viewed as a direct representation of reality. The latter was found to be the most frequently made error. Wemyss and van Kampen [3] categorized interpretations of numerical linear distance time and $y(x)$ graphs from first-year university students enrolled in an algebra-based course. They found that both context and prior learning are likely of influence to explain poor performance and that the ability to determine the slope of an $y(x)$ graph and a correct qualitative understanding of distance-time graphs is insufficient for correctly determining the speed in a distance-time graph. Of note here is that they explicitly asked students to determine the slope or the speed at a particular instant. In their pretest data, they found that just under 20% of all students could determine the speed from an $x(t)$ graph, and that just over 50% could determine the slope in an $y(x)$ graph. Slope or height confusion and iconic interpretations were only present in a small minority of their data, whereas interval or point confusions occurred with 53% and 21% of their students in the kinematics and the context-free questions, respectively. When students were asked a similar question but with a graph in the context of water level versus time, their accuracy doubled compared to the distance versus time questions. Furthermore, almost twice the percentage of students used interval reasoning compared to the percentage for the distance-time questions. A follow-up study was performed by Bollen *et al.* [7]. They adapted the categorization and administered a similar test with first-year university students in calculus-based courses in the Basque country (Spain) and in Belgium. Their results showed similar difficulties, often related to dividing two coordinates to calculate the speed. Furthermore, they found that the context influences the success rate and that qualitative understanding of kinematics is important but not sufficient to determine the speed in $x(t)$ graphs. The study by Planinic *et al.* [2] investigated students' understanding of linear graphs in mathematics, kinematics, and contexts other than physics. In their multiple choice test, students can choose between three comparative statements about the slope or the area under a graph, and must select the correct one. The results showed that mathematics is the easiest context for students and the other two are on par in difficulty. Additionally, determining the slope was found to be more difficult compared to determining the area under a graph— the other concept they studied. Concerning the y intercept, Davis [1] showed that students naturally and successfully start to use the informal terminology "starting point" when confronted with real-world contexts in multiple representations before the introduction of formal mathematics terminology. The study showed that a disconnect between formal and informal terminology in teaching activities results in a fragile understanding of the y intercept, and strongly suggests to include explicit learning activities to connect these terminologies. After the introduction of formal mathematics terminology, students were confronted with abstract tasks (no real-world context) using the same representations and informal terminology which the researchers found to negatively impact students' performance. Adu-Gyamfi and Bossé [5] showed that even students who perform admirably at representation-related tasks concerning functions might still have limited mathematical understanding. This implies that testing students' understanding of functions should likely include, but certainly not be limited to, various

representations. Also De Bock *et al.* [6] found that the representational format significantly influenced student performance as a main effect and as an interaction effect with the specific type of function (proportional, inverse proportional, affine with negative slope, and affine with positive slope). Ibrahim and Rebello [21] studied students' strategies when solving tasks concerning kinematics and work. They showed that, irrespective of the various representations offered in the questions' format (verbal, graphical, and symbolic), students preferred manipulating equations and that they more often rely on equations to solve kinematics tasks whereas for tasks related to work they preferred a qualitative approach, thus indicating the dependency of the solution strategy on the context. Also Acevedo Nistal, Van Dooren, and Verschaffel [18] came to the conclusion that their students in secondary education prefer to use formulas over tables or both to solve for the dependent or independent variable in linear function problems. They compared students' choices and performance in choice and no-choice conditions for the representation in which to solve the problem. Additionally, they found that the choice to use formulas, or the pressure students feel to use formulas, increases with grade. The representation in which a problem is presented and solved can strongly influence students' performance and the use of representations has been found to be very different between novices and experts [23]. Representational fluency in particular has been of interest to the research community, and notably the Representational Fluency Survey (RFS) [24] has been developed and applied [25] to assess this ability with university students in the context of physics. In light of this, we previously developed a multiple choice test [8] for students' representational fluency—specifically for the ability to translate between representations—of linear function problems in mathematics and kinematics. Our results showed that significant main effects of the representational transition, the function type (signs of slope and *y* intercept), and the context (kinematics and mathematics) exist. The study we present here is complementary to our previous one: now we focus more on qualitative insights and on the use of strategies and frequent errors.

All these findings point to a disconnect between contexts, a strong influence of, and a disconnect between representations, and specific difficulties depending on the task and the concepts related to linear functions. This strongly suggests that students' understanding is compartmentalized to the specific situation in which they first learned about a specific topic. This is not unusual in the early stages of learning, but the referenced studies in higher education show that this compartmentalization continues to exist. To amend the situation, links should be actively constructed to connect these islands of understanding and achieve deeper understanding with the flexibility to apply knowledge and skills to new situations and achieve an efficient transfer thereof. To do so, it is important to understand students' difficulties that hamper the construction of these links. More than anything, the literature shows that these factors (contexts, representational format, function type, and task) should be included when assessing students' understanding.

III. RESEARCH DESIGN

A. Research questions

The broad goal of our study is to gain more insight in students' understanding of linear relations in kinematics [$x(t)$] and in mathematics [$y(x)$]. Since effects from context, representational format, function type, and task have been observed in literature, we include these factors in our research and ascertain their effects in a quantitative analysis followed by a qualitative analysis to gain as much insight as possible in students' strategies and errors. Our specific research questions are as follows:

(1) How do the accuracy and strategy of grade 9 students compare between solving $x(t)$ problems in kinematics and solving isomorphic $y(x)$ problems in mathematics?

(2) How do grade 9 students interpret negative velocity in $x(t)$ graphs and in algebraic formulas; and how does this compare to positive velocity and to negative and positive slope in the isomorphic $y(x)$ questions in mathematics?

(3) How does the use of graphs or algebraic formulas affect grade 9 students' accuracies and solution strategies when determining or comparing velocity or initial position in linear $x(t)$ problems in kinematics and slope or y intercept in linear $y(x)$ problems in mathematics?

(4) How can we categorize grade 9 students' strategies when comparing and identifying velocity or initial position in linear $x(t)$ problems in kinematics and slope or y intercept in linear $y(x)$ problems in mathematics?

To answer these questions we design and validate a test in the next sections. An English translation of the resulting test is available for download as Supplemental Material [26].

B. Design choices

A linear relation can be characterized by the combination of the signs of the y intercept and the slope. To limit the number of items, we select the two function types most familiar to our respondents: positive y intercept and positive slope, and positive y intercept and negative slope; each of these is represented by either a graph or an algebraic formula. The subject of each question is either of the two concepts: the y intercept or the slope. In addition, two kinds of tasks are used: determine the value of a concept or compare the value of a concept between two different situations. We combine the representations and the tasks into three different question types: determine via a graph, determine via a formula and

compare via a graph. The other combinations are omitted because they are deemed less prone to errors and to limit the number of items. The contexts under investigation are kinematics and mathematics. In kinematics, students are presented with a description of one or two vehicles driving along a straight road with constant velocity, thus performing a 1D uniform linear motion. In mathematics, the questions concern a function $f(x)$ and no additional context is provided. To gain insight in student reasoning, we choose an open-ended question format including the explicit requirement for an explanation in some strategically selected questions. Finally, to obtain a consistently structured test and clearly defined questions, some additional design choices are made which are illustrated in the examples provided in Fig. 2:

  (i) In graphs, we focus on the first quadrant because we want to avoid issues which the use of other quadrants might trigger. We do show small sections of the other quadrants and we draw the curves in those quadrants depending on our students' familiarity with negative values for the respective variables. The result is that we include negative values for position, but exclude negative values for time. In mathematics we include negative values for all variables.

 (ii) In comparison questions with graphs, each curve is drawn with a different thickness to better distinguish them. Furthermore, the two curves intersect in the first quadrant and each curve is labeled twice: once at the left and once at the right of the intersection point. These choices are motivated by the results from a pilot study which showed some confusing answers in which students referred to "the highest curve" or "the lowest curve" which makes interpretation of their explanations inconsistent.

(iii) In kinematics, units for the time t—always in seconds—and for position x—always in meters—are provided in the textual description of the situation and in the case of a graph they are also shown on the axis labels; coefficients in the formula representation do not include units.

(iv) Formulas in kinematics use the $x = x_0 + vt$ format in which the term with the lowest degree comes first, a format more often used in physics. In mathematics, formulas use the $f(x) = ax + b$ format in which the term with the highest degree comes first, which is more common in mathematics.

In addition, there is an important issue to highlight: a stationary object must accelerate before achieving a constant nonzero velocity. In the graphical kinematics items, we implicitly assume that the measurement starts at $t = 0$ and that the motion started before that moment, thus avoiding the acceleration issue.

C. Test structure

The design choices result in a 24-item test structured as shown in Fig. 1. Figure 2 shows some examples. The test


FIG. 1. Test structure: two signs for the slope (positive and negative); two concepts (slope and y intercept); three question types using two tasks (compare or determine) and two representations (graph and formula); two contexts [kinematics (K), mathematics (M)]. The items marked in gray require an explanation.

consists of two as isomorphic as possible context blocks: 12 kinematics items and 12 mathematics items. To avoid possible learning effects of solving a particular context block first, half of the respondents first complete the kinematics part and then the mathematics part, and the other half has the order reversed. The order of the items within each block is also randomized. Both linear function types under study have a positive y intercept but can have a positive or a negative slope. To distinguish between these we will simply refer to their slope sign. Each sign is questioned 6 times per context. Each item has one concept as its subject: the slope or the y intercept in mathematics and the velocity or initial position in kinematics. Furthermore, there are three different question types. In the first type, students must compare the slope or y intercept of two linear relations represented in a single graph. In the second, they must determine slope or y intercept given a graph. In the third question type, they must determine slope or y intercept given an algebraic formula. For the majority of the items the respondents are required to provide an explanation for their answer, using the additional instruction "*Explain your answer.*" This was not done for all items to limit the test time. The items which require an explanation are marked in Fig. 1 and are strategically chosen to gain as much understanding as possible, to avoid repetition of similar answers, and to collect answers about the items for which we expected the most difficulties, e.g., the comparison of negative velocities in a graph.

We designed this test to evaluate whether students can interpret concepts related to linear relations in different contexts. The test is not a concept test or an inventory [27] in the sense that it aims to measure students conceptual understanding in a particular domain, but rather aims at comparing students' understanding of particular isomorphic concepts between different domains like physics or mathematics. Its structure is in line with many other existing tests. It is most similar to the ones from Bollen *et al.* [7] and Wemyss and van Kampen [3] which study a similar topic in isomorphic contexts and make use of written explanations for analysis. Also comparable are the test from Planinic *et al.* [2,4] and Ivanjek *et al.* [28], who studied line graphs in various contexts using multiple choice tests and tests with open questions requiring an

K6

A train is moving along a straight track. The graph shows the position x of the train as function of the time t. The position is in meter, the time in seconds.
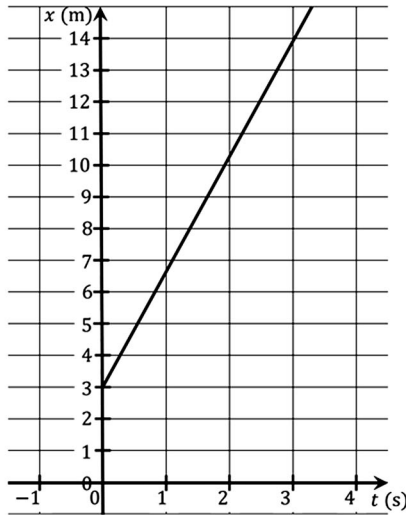
Determine the velocity of the train. Explain your answer.

K11

A car is moving along a straight road. The position x of the car as function of the time t is given by

$$x = 5 - 10t.$$

The position is in meter, the time in seconds.
Determine the initial position of the car. Explain your answer.

K4

Two cyclists a and b are riding along a straight bicycle path. The graph shows their position x as function of the time t. The position is in meter, the time in seconds.
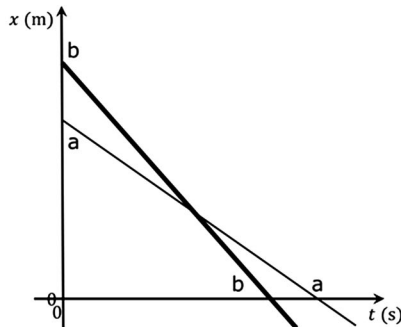
Which cyclist has the largest velocity? Mark the correct answer and explain.

| **Cyclist a** | **Cyclist b** | **Cyclists a and b have the same velocity** |

M6

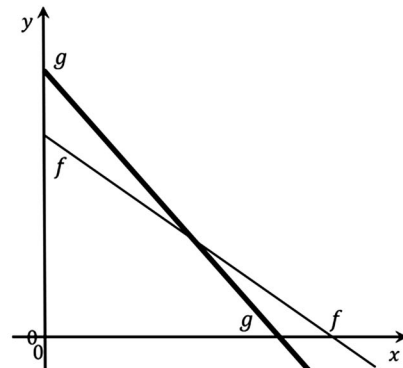The figure shows the graph a function f.
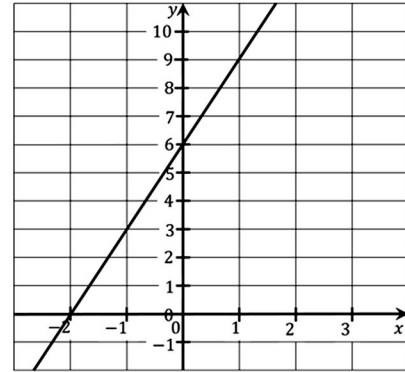
Determine the slope. Explain your answer.

M11

A function f is given by

$$f(x) = -8x + 2.$$

Determine the function value of 0. Explain your answer.

M4

The figure shows the graphs of two functions f and g.

Which of these functions has the largest slope? Mark the correct answer and explain.

| **f** | **g** | **f and g have the same slope** |

FIG. 2.   Three kinematics (K) items and the three isomorphic mathematics (M) examples illustrating the different factors in the test design.

explanation. In addition, Bassok and Holyoak [29] compared constant acceleration kinematics and the isomorphic algebra question in a textual representation using think-aloud protocols. Furthermore, similar test structures have also been used for different topics by, e.g., Pollock *et al.* [30] when they compared the understanding of integrals in physics and mathematics; as well as by Barniol and Zavala [31] who studied the effect of context on students' understanding of vector concepts in their TUV-12 test and related

studies [32], though the latter made use of multiple-choice questions.

IV. TEST ADMINISTRATION

A. Participants

The test was administered in 2017 in 17 classes distributed over 7 secondary schools in Flanders (Belgium) with one physics teacher and one mathematics teacher per school

for these classes. A total of 253 students with ages 14–15 participated in the study. These grade 9 students were enrolled in programs with a strong science component. Our cohort consists of 117 females and 136 males. All of them had received instruction on linear relations in mathematics and on basic 1D kinematics in physics in compliance with government imposed mandatory learning goals. Typically, instruction starts with examples from everyday life and simple experiments, followed by the introduction of the mathematical formula for average velocity, after which the graphical representation and interpretation in $x(t)$ graphs for uniform linear motions and other linear motions is discussed, and finally, the necessity to use vectors and the vector properties of velocity are made clear through examples. Additionally, experiments such as electric toy cars tracked by a motion sensor which outputs the measurement data in a graph are strongly encouraged. There is usually little to no focus on the algebraic formulation of a uniform linear motion in the physics course at this point. Total time allocated for this is 6 teaching hours of 50 minutes each. In mathematics students learn about linear relations and linear functions in graphical representations, algebraic representations, and tabular representations during 17 teaching hours of 50 min each. These lesson series start with the definition of the linear function, drawing linear graphs, translating between the various representations, discussing the meaning of the coefficients, studying the signs of the coefficients, of the dependent and independent variable, and solving standard exercises. Six schools participating in this study were part of a larger research project for which they had agreed to allow testing to take place within the classes with the required profile. One additional school was also contacted and voluntarily agreed to have classes with the requested profile to participate in the study. The locations of these schools are distributed across Flanders and all of them provide "aso education" in which students receive education on a broad number of topics to prepare them for higher education. All students in each class had to participate, but parents were given the possibility to have their child excluded from the test for, e.g., performance anxiety reasons, which only occurred a negligible number of times. Ethical approval for this study was granted by the Sociaal-Maatschappelijke Ethische Commissie (SMEC) Social and the Societal Ethics Committee of KU Leuven.

B. Test procedure

The test is administered as a paper-and-pencil test to allow the respondents as much freedom as possible in the way they construct and explain their answers. The students are notified beforehand that a test will take place as part of a research project, that they do not need to prepare for it and that no credit can be earned. At the start, they are informed that the available time is 45 min, that it is crucial to explain their answer whenever this is required, and that they may use whatever method they prefer in their explanation, e.g.,

write a text, make a calculation, make a drawing, annotate the graph or formula, etc. Furthermore, if they are unable to answer a question they are asked to mark that question with a forward slash to indicate that they have read the question but do not know the answer.

C. Data analysis

In the quantitative analysis, coding of the data is dichotomous: unanswered and incorrect items are coded as incorrect (i.e., accuracy 0). An answer is considered correct if both the answer and the explanation (when required) are correct (i.e., accuracy 1). When the answer is correct but the explanation is not, the item is marked as incorrect. When the answer is correct but the required explanation is missing, the item is marked correct, giving the respondent the benefit of the doubt. For questions in which no explanation is required the answer is coded as 1 when correct and as 0 when incorrect. Item K4 (included in Fig. 2) requires special attention. This item asks to compare two negative velocities in a graph, which can create a polarizing discussion in Dutch since there is no linguistic distinction between "velocity" and "speed," both are translated as "snelheid." This issue has also been discussed in our previous work [8]. In this study we choose to interpret snelheid as velocity in all questions, meaning that we expect students to take both the magnitude and the direction into account. Essentially this means that, since these are all 1D situations, we inquire about the scalar component v_x of velocity $\bar{v}$ whenever we use snelheid in this test. This choice means that we consider "Cyclist a" to be the correct answer for question K4, which is in line with the isomorphic question M4 (also included in Fig. 2), in which answer "f" is correct. Also in questions with an algebraic formula, the minus sign must be included in the answer, e.g., in K11 in Fig. 2 the correct answer is -10 m/s.

Analysis is done using generalized estimating equations (GEE) [33,34], which is an adaptation of logistic regression analysis and takes the repeated measurements of our dichotomous scoring into account [35]. The IBM ® SPSS ® Statistics version 24.0.0.0 software package is used. Settings in the GEE functionality are binomial probability distribution with the logit link function, the Fisher parameter estimation method, the type III error analysis type, Wald statistics, and the Bonferroni correction to adjust for multiple comparisons. Our statistical model includes *context* (1D kinematics or mathematics), *QType* (compare given a graph, determine given a graph, determine given a formula), *concept* (y intercept, slope) and *slope sign* (+, −), *gender* (female, male). We include gender since this has been shown to have a possible (interaction) effect with the external representation [36], which is related to the question type in this case. All main effects, two-way effects, and three-way effects are included in the model.

TABLE I.   Validation results sorted by *context* (indicated by "K" for kinematics and "M" for mathematics in the question number Q); by *representation* (graph or formula); by *task* (compare or determine), by *slope sign* ("+" for positive, "−" for negative) and by *concept* (y intercept or slope). The statistical parameters are item difficulty index P and its standard deviation SD; Cronbach's α when the item in question would be deleted; and item discrimination index D. Items marked with ∗ require the respondent to provide an explanation for their answer.

Q	Representation	Task	Slope sign	Concept	P	SD	α when deleted	D
K1	Graph	Compare	+	y intercept	0.88	0.33	0.783	0.34
K2∗	Graph	Compare	+	Slope	0.92	0.28	0.790	0.18
K3	Graph	Compare	−	y intercept	0.83	0.38	0.788	0.26
K4∗	Graph	Compare	−	Slope	0.13	0.34	0.795	0.12
K5	Graph	Determine	+	y intercept	0.74	0.44	0.783	0.46
K6∗	Graph	Determine	+	Slope	0.48	0.50	0.794	0.32
K7∗	Graph	Determine	−	y intercept	0.66	0.47	0.782	0.53
K8∗	Graph	Determine	−	Slope	0.14	0.35	0.791	0.15
K9	Formula	Determine	+	y intercept	0.37	0.48	0.779	0.53
K10∗	Formula	Determine	+	Slope	0.11	0.32	0.784	0.26
K11∗	Formula	Determine	−	y intercept	0.35	0.48	0.781	0.51
K12∗	Formula	Determine	−	Slope	0.05	0.21	0.791	0.07
M1	Graph	Compare	+	y intercept	0.53	0.50	0.779	0.66
M2∗	Graph	Compare	+	Slope	0.85	0.36	0.786	0.28
M3	Graph	Compare	−	y intercept	0.54	0.50	0.786	0.49
M4∗	Graph	Compare	−	Slope	0.41	0.49	0.791	0.44
M5	Graph	Determine	+	y intercept	0.39	0.49	0.776	0.66
M6∗	Graph	Determine	+	Slope	0.70	0.46	0.780	0.54
M7∗	Graph	Determine	−	y intercept	0.37	0.48	0.770	0.82
M8∗	Graph	Determine	−	Slope	0.67	0.47	0.787	0.43
M9	Formula	Determine	+	y intercept	0.36	0.48	0.777	0.65
M10∗	Formula	Determine	+	Slope	0.74	0.44	0.783	0.51
M11∗	Formula	Determine	−	y intercept	0.36	0.48	0.772	0.74
M12∗	Formula	Determine	−	Slope	0.71	0.45	0.780	0.62

For the qualitative analysis, we use open coding to identify common strategies and recurring ideas across different students. This results in a first categorization scheme describing student explanations. To achieve a sound categorization, the scheme is—based on a subset of the data—refined by a second researcher and then independently used to analyze a new subset of the data by the first author and a colleague. The final categorization scheme is discussed in Sec. VI.

V. VALIDATION AND STUDENTS' ACCURACY

In this section the validation of the test is discussed, followed by a quantitative analysis of the answers.

A. Validation

Here, we report on key validation parameters, which show that the test has good internal consistency, a wide range of item difficulty indexes and item discrimination indexes, and a good discriminatory power which makes the test well suited for our goals. An overview of these results on a per-item basis is provided in Table I.

1. Internal consistency

The internal consistency is assessed using Cronbach's α, which is $\alpha = 0.791$ and is an acceptable value just shy of 0.80, the criterion for a "fairly high" internal consistency [37]. Since omitting certain items from the test might increase this value, α was also calculated for each case in which an item i is omitted from the test. These values are shown in Table I and show that omitting a test item would only marginally increase the internal consistency by a maximum of 0.04 in the case of item M4.

2. Item difficulty index

Because of our dichotomous coding for the accuracy, the average accuracy of each item translates directly into the item difficulty index P:

$$P = \frac{N_c}{N},\qquad(1)$$

in which N_c is the number of answers with accuracy 1, and N is the total number of answers, which is the number of participants: $N = 253$. A higher value for P is indicative of

an easier item and a lower value indicates a more difficult question. Table I shows an item difficulty range between 0.05 and 0.92, and an average item difficulty of 0.51 ± 0.26.

The items with the highest average scores are K1, K2, K3, and M2, which are all "compare given a graph" questions, three out of four have a positive slope, and three out of four are kinematics questions. This illustrates the contrast between the contexts since the most difficult questions, as well as the majority of the easiest questions are found in kinematics, while the mathematics items do not have such extreme item difficulty indexes and are less broadly distributed. Furthermore, these observations highlight the important effect of the slope sign on the average accuracy and the difficulty students have with negative slopes in kinematics.

Items K4, K8, K10, and K12 stand out because they have very low item difficulty indexes accompanied with some of the lowest standard deviations, meaning that these items are very difficult for the large majority of students. The common factor for these four questions is that all are kinematics items in which the concept is the velocity. The sign of the velocity is negative for K4, K8, and K12; and the task in items K8, K10, and K12 is to determine the velocity. The counterparts of these questions in mathematics (M4, M8, M10, M12) show no such pattern and the counterparts which query about the y intercept (K3, K7, K9, K11) do not show any similarities either. Furthermore it is clear from the lower item difficulty indexes of K9 and K11 that determining the initial position given a formula is far more difficult than when presented with a graph.

Of note here is the extreme contrast between item K2 and K4, which are analogous questions except for the sign of the velocity.

Another remarkable difference is clear when comparing the K5–K8 block with the isomorphic M5–M8 block, i.e., the items in which to determine the y intercept or slope given a graph. For these blocks, the item difficulty indexes for the kinematics items have the opposite structure of those in mathematics: it is easier to determine the initial position than the velocity in kinematics graphs compared to mathematics, where it is easier to determine the slope than the y intercept. Part of this might be explained by the lower familiarity students have with the term "function value" in the wording "the function value of 0" in the mathematics items—compared to the more familiar "initial position" in kinematics. Although the wording is part of the official curriculum, many students made note of this lack of familiarity during the test or in their explanations. This is also clear in the items in which the y intercept has to be compared (K1, K3, M1, M3). For these items we did not require an explanation because we expected them to be easy and result in fewer interesting explanations. This proves to be correct in kinematics with accuracies for K1 and K3 above 0.80, but was not the case in mathematics

(M1 and M3) which have an accuracy of 0.53 and 0.54, respectively. This difference between the isomorphic items indicates that students did not recognize the similarities between the contexts.

3. Discriminatory power

The discriminatory power of a test is a measure of its ability to discriminate between respondents with high and low ability. On a per-item basis, we can calculate the item discrimination index $D = PU - PL$, in which PU is the item difficulty index of the upper group of test scores, i.e., the 27% of highest scoring students; similarly, PL is the item difficulty index of the lower group, i.e., the group with the 27% lowest scoring students [37]. The higher the value of D for an item, the better that item discriminates. The results are shown in Table I and range between 0.07 and 0.82 with an average of 0.44 ± 0.20. Again the same four kinematics items (K4, K8, K10, and K12) stand out for their low score; which—combined with their lower item difficulty index—shows that the large majority of students has severe difficulties with these questions and that the score distribution for these four items is very narrow. Additionally, item K2—a kinematics item to compare positive velocities in a graph—has a very low item discrimination index in combination with a very high item difficulty index, indicating that this item is very easy for the large majority of students and that its score distribution is very narrow.

To assess the discriminatory power of the full test we use Ferguson's delta δ, which is the ratio of the interperson differences to the maximum number possible:

$$\delta = \frac{(k+1)(n^2 - \Sigma_{i=1}^k f_i^2)}{kn^2}, \qquad (2)$$

with $k = 24$ the number of items in the test, $n = 235$ the sample size, and f the frequency of a certain score i, with i from 1 to k. This results in $\delta = 0.971$, which easily satisfies the minimum required value of 0.9, meaning that the test has good discriminatory power [37].

B. Main effects and interaction effects

In this section we discuss the results from the GEE analysis, the errors in this section are always standard errors. The GEE analysis shows four significant main effects. The first is *context*, for which we find Wald $\chi^2(1) = 27.363$ and $p < 0.001$ and a significant difference ($p < 0.001$) between a mean of 0.45 ± 0.02 for kinematics and 0.56 ± 0.02 for mathematics, which is in line with the results from our previous study [8]. The second significant main effect is that from *QType* with Wald $\chi^2(2) = 194.836$ and $p < 0.001$. The means are 0.67 ± 0.02 for "compare given a graph," 0.51 ± 0.02 for "determine given a graph," 0.33 ± 0.02 for "determine given a formula," and each pairwise comparison is significant ($p < 0.001$). Third, for

TABLE II. Significant two-way and three-way interaction effects with Wald χ^2, degrees of freedom (DOF), and significance (p value) up to three digits.

Term	χ^2	DOF	p
*Context * concept*	348.240	1	0.000
*Context * slope sign*	84.922	1	0.000
*Context * QType*	160.690	2	0.000
*QType * slope sign*	117.569	2	0.000
*Concept * slope sign*	156.080	1	0.000
*Concept * gender*	6.382	1	0.012
*Context * QType * Concept*	22.178	2	0.000
*Context * QType * slope sign*	10.351	2	0.006
*Context * concept * slope sign*	45.961	1	0.000
*Context * slope * gender*	4.009	1	0.045
*QType * concept * slope sign*	91.056	2	0.000

FIG. 4. GEE results for the term *context * slope sign*. The p value of the comparisons is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. A highly significant difference is found between the contexts for negative slopes.

concept the results are Wald $\chi^2(1) = 17.567$ and $p < 0.001$ with a significant ($p < 0.001$) difference between 0.55 ± 0.02 for y intercept and 0.46 ± 0.02 for slope. For *slope sign* this results in Wald $\chi^2(1) = 227.176$ and $p < 0.001$. Pairwise comparison of the slope sign shows a significant difference ($p < 0.001$) with a mean of 0.61 ± 0.02 for positive slope and 0.40 ± 0.02 for negative slope. Lastly, the variable *gender*—just like in our previous study [8]—does not result in a significant main effect: Wald $\chi^2(1) = 1.356$ with $p = 0.244$.

There are also many significant interaction effects, which are shown in Table II.

First, the term "*context * concept*" for which the results are provided in Fig. 3. This shows that, concerning the y intercept, students perform significantly ($p < 0.001$) better in kinematics than in mathematics, while the opposite is true for questions concerning the slope.

The results for the term "*context * slope sign*" are shown in Fig. 4, which show that students' performance for questions with positive slope is very similar in kinematics and mathematics. For negative slope though, there is a
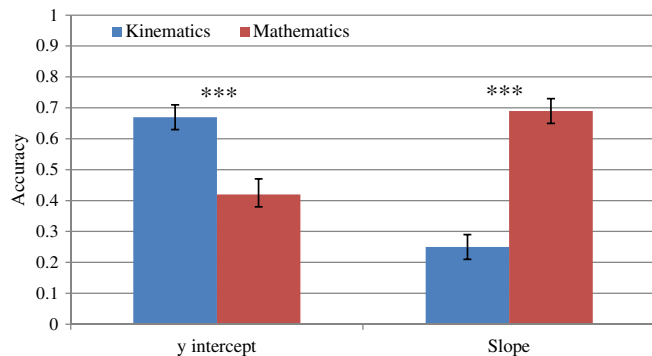
significant ($p < 0.001$) difference between the two contexts which indicates that students have far more difficulties with negative velocity than they do with negative slope. The majority of this difference originates from questions in which the velocity or slope is the concept and the ∼50% drop between positive and negative scores in kinematics happens equally across all three questions types.

Figure 5 shows the results for the term *context * QType*, which illustrates that there are significant differences ($p < 0.001$) between the two contexts for "determine via formula" and "compare via graph," in which the former has the largest difference between the contexts. Furthermore, there are no significant differences among question types in mathematics, but very significant differences ($p < 0.001$) between all of them in kinematics.

Since the slope causes many difficulties in kinematics, it is interesting to zoom in on the results for the slope in the term *context * QType * concept*, which are shown in Fig. 6. This graph shows significant ($p < 0.001$) differences between contexts for the two question types including the "determine" task. Furthermore, it shows no significant differences between question types in mathematics, but

FIG. 3. GEE results for the term *context * concept*. The p value of the comparisons is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Highly significant differences are found between the contexts for each of the concepts.

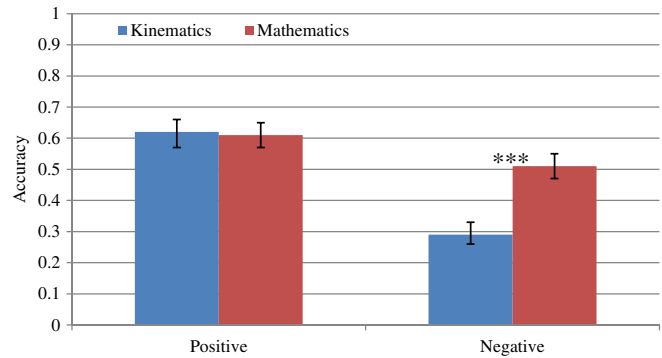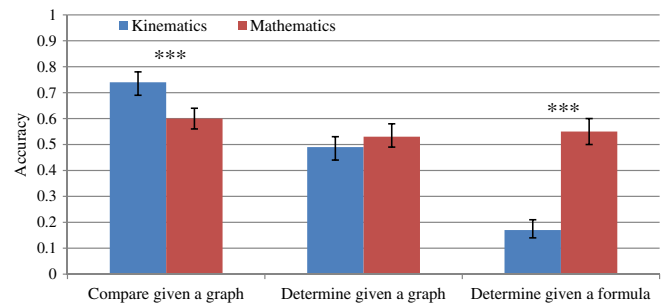FIG. 5. GEE results for the term *context * QType*. The p value of the comparisons is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Highly significant differences are found between the contexts for compare given a graph and for determine given a formula question types.
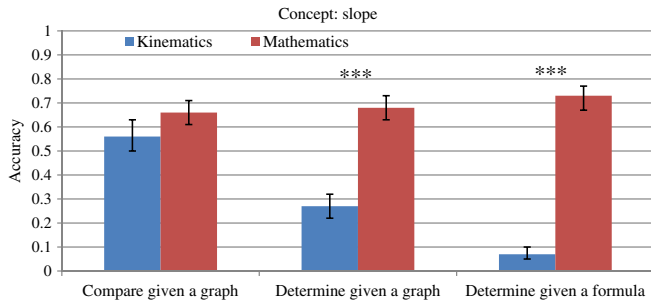
FIG. 6.   GEE results for the slope in the term *context * QType * concept*. The p value of the comparisons is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Highly significant differences are found between the contexts for determine given a graph and for determine given a formula question types.

very significant ($p < 0.001$) ones in kinematics. The highest mean accuracies in kinematics are achieved in compare via graph because the very high accuracies for positive slope questions in this question type compensate the very low accuracies for negative slope questions in this question type, thus averaging out slightly above 50%. For the other two question types though, the mean accuracies in kinematics are very low for both the negative and positive slopes resulting in significantly lower mean accuracies overall.

Figure 7 shows the significant difference ($p < 0.001$) between physics and mathematics and positive and negative slope for questions concerning the slope. Most importantly, it shows that negative slope is far more difficult in kinematics compared to mathematics.

These results support those from the validation, which is that students have the most difficulties with items in kinematics, concerning slope, with negative velocity in all three question types, i.e., items K4, K8, and K12. In addition item K10, in which students must determine the positive slope when given a formula, also poses increased difficulties for students. In mathematics the main difficulty
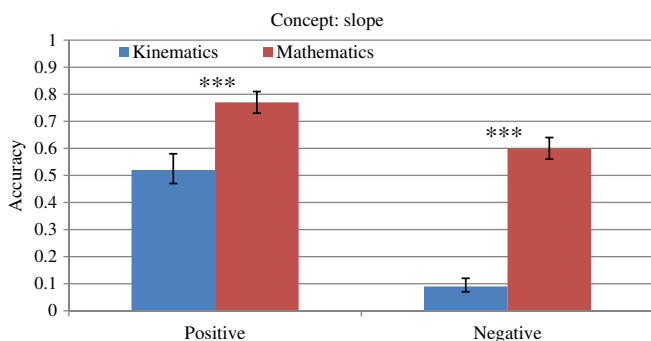

FIG. 7.   GEE results for the slope in the term *context * concept * slope sign*. The p value of the comparisons is indicated by * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Highly significant differences are found for the slope between the contexts for both slope signs.

is determining the y intercept in both graphs and algebraic formulas. This is possibly explained by the relative unfamiliarity with the terminology "function value" in the wording "the function value of 0" to indicate the y intercept in the mathematics items. Furthermore, there is a larger focus on the x intercept than there is on the y intercept in mathematics lessons. In kinematics the initial position of a motion is of greater importance since in this context the choice of the coordinate system and its origin is often intelligently chosen so that the initial position is in the origin. Furthermore, terminology such as "initial position" is more intuitively clear for students compared to the function value of 0 as was also the case in Davis' study [1].

VI. QUALITATIVE RESULTS

For the qualitative analysis, a categorization scheme was constructed bottom up from the raw data. First, a pilot study with 181 students and a first version of the test was performed and a first version of the scheme was constructed. A second researcher made refinements to the test to make the graphs and the wording clearer and to create a better distinction between the two contexts, but the structure of the test remained. The scheme was optimized by grouping some categories and creating a better distinction between others. That optimized scheme was then used to categorize a subset ($n = 42$) from the new data from the refined test by a third researcher, which resulted in some additional refinements. The final scheme was then used to categorize all the new data ($n = 253$) by the second researcher. In this section we describe the scheme, validate its reliability, and discuss trends in the data.

A. Categorization scheme

The final categorization scheme is shown in Table III. It is split by concept (y intercept or initial position and slope or velocity). The first seven categories are shared between both concepts followed by two and four concept specific categories. Answers usually fall in only a single category, but a non-negligible number of questions had answers which made use of multiple strategies, in which case they were put into all suitable categories.

  I1/S1: Location in an equation: A coefficient in the equation is marked by underlining it or encircling it. Alternatively, the student explicitly writes that, e.g., "6" is the answer because it is located before the "x."

  I2/S2: Identification of the x intercept: Instead of identifying the y intercept or slope they identify the x intercept (graphical or symbolic).

  I3/S3: Construct an equation: Change of representation. When presented with a graph, an equation was constructed.

  I4/S4: Construct a graph: Change of representation. When presented with an equation, a graph was

TABLE III.   Bottom-up constructed categorization scheme.

y intercept or initial position	Slope or velocity
I1: Location in an equation	S1: Location in an equation
I2: Identification of the x intercept	S2: Identification of the x intercept
I3: Construct an equation	S3: Construct an equation
I4: Construct a graph	S4: Construct a graph
I5: Construct a table	S5: Construct a table
I6: Intersection with vertical axis	S6: Reasoning with "steepness"
I7: Calculate a specific function value	S7: Drawing a triangle on a graph
	S8: Ratio of differences
	S9: Ratio of coordinates
I99: Other	S99: Other

constructed. This need not be a detailed one, a simple sketch suffices.

I5/S5: Construct a table: Change of representation. A table is constructed with two or more data points.

I6: Intersection with vertical axis: When a graph is given or constructed, the y intercept is marked by, e.g., encircling it or drawing an arrow towards it. Often this is accompanied by written explanations such as "This is where the line/graph/cyclist starts/begins."

I7: Calculate a specific function value: When an equation is given or constructed, the value of the dependent variable is calculated for a specific value of the independent variable, often with $x = 0$ or $x = 1$.

S6: Reasoning with "steepness": Qualitative reasoning concerning the steepness of the graph. Often wording similar to "Graph a is steeper than graph b" is used.

S7: Drawing a triangle on a graph: Drawing a triangle with a vertical and a horizontal side under or above the line of a graph to indicate the step change in both directions. Often these triangles are accompanied by the change in value along each axis with notations such as "+1" and "-2." The size of the triangle does not matter.

S8: Ratio of differences: Calculation of the slope using $\Delta y / \Delta x$ or $(y_2 - y_1)/(x_2 - x_1)$, or written explanations such as "In 1s the cyclist traveled 2m;" "If x increases with 1, then y increases with 3;" "If you go one unit to the right, then you go 3 units up." A specific case in this category is when the ratio of the y value of the y intercept over the x value of the x intercept is calculated, which is essentially the ratio of the differences from the intercepts with respect to the origin. More qualitative explanations also fit in this category, such as: "In the same time, cyclist b traveled more meters;" "Cyclist b starts later, but still overtakes cyclist a."

S9: Ratio of coordinates: The ratio y/x of some coordinate (x, y) is calculated. This includes cases in which the correct formula is written, e.g., $v = \Delta x / \Delta t$,

but actually $v = x/t$ is calculated, signaling the disregard or the misunderstanding of Δ in the formula. Often $f(1)/1$ or $f(x = 1)$ is calculated. Also written explanations such as "The y value of function f is larger than that of function g for the same x value" with comparison questions are in this category.

I99/S99: Other: All answers which are less frequent and do not fit in any of the other categories. For example, confusion of y intercept and slope; calculations with the coefficients of a formula such as $a * b$ or $a + b$ in $f(x) = ax + b$; identifying the slope as $6x$ in $f(x) = 6x + 2$; in the case of a y intercept question, providing coordinates such as $(0,2)$ given the equation $f(x) = 6x + 2$, which is always considered incorrect since the questions ask for the function value or the initial position; or in case of a y intercept question, providing a coordinate constructed of the x intercept and the y intercept such as $(-\frac{1}{3}, 2)$ for $f(x) = 6x + 2$.

Categories I3/S3, I4/S4, and I5/S5 all describe a change in representation, these categories are always accompanied by another strategy to fully answer the question. We deliberately keep these categories separate so we can compare their prevalences. Categories S7 and S8 are both essentially a ratio of differences, but our data strongly suggest that it is useful to distinguish the two.

B. Interrater reliability

To validate the classification scheme, the interrater reliability is calculated using Cohen's κ, which provides a measure for the agreement between two raters of the same data. The coefficient can be calculated as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \qquad (3)$$

in which p_0 is the accuracy of the agreement between the raters, i.e., the percentage of agreement; and p_e is the probability of agreement by change. We found a Cohen's κ of $\kappa = 0.607$ and calculation of κ on a per item basis showed values often much higher than that, which indicates good agreement [38,39] between the categorization of the second and third researcher for a subset of the data ($n = 42$).

C. Results

In this section we describe students' use of the strategies in the categorization scheme. Since not all questions in the test require an explanation and since not every answer contained an explanation (required or not), we express frequencies relative to the number of answered questions for which at least one strategy was discerned by the researcher. This means that unanswered questions are not included. With this definition, the maximum frequency of 100% for a certain category means that every answered question was provided with an explanation and that this

TABLE IV. Frequencies (in %) for the y intercept questions. f is across both contexts, f_K is in the kinematics context and f_M is in the mathematics context. The percentages are relative to 1600, 723, and 877, respectively.

Category	f	f_K	f_M
I1: Location in an equation	4.1	4.7	3.6
I2: Identification of the x intercept	27.1	9.7	41.5
I3: Construct an equation	12.9	4.1	20.2
I4: Construct a graph	2.9	5.0	1.1
I5: Construct a table	1.4	1.4	1.5
I6: Intersection with vertical axis	24.3	39.6	11.7
I7: Calculate a specific function value	23.6	17.6	28.6
I99: Other	17.3	24.8	11.2

category was used in all those explanations. To clarify this a bit more, let us assume that the data contains 2000 questions which were answered *and* for which an explanation was provided which is categorized in *at least* one category. Since some explanations contain multiple categories, the total number of items in all categories is higher than 2000, but 2000 is the maximum prevalence for each single category, i.e., this prevalence would result in a frequency of 100%. If 400 out of 2000 cases with *at least* one category are categorized in category X, then the frequency for X is 20%. A discussion about the highlights will be provided and the main discussion will be done in Sec. VII.

1. y intercept

The results for the y intercept questions are shown in Table IV. Answers without any explanation are omitted from this analysis. When an explanation is provided but not requested, the answer is included in the data. This results in far more data than anticipated than from the four questions which explicitly require an explanation for the y intercept (shown in Fig. 1). The resulting data set contains 1600 uses of *at least* one strategy across the *twelve* questions about the y intercept, 723 in kinematics and 877 in mathematics. The gathered data structure is largely in line with the full test structure in Fig. 1. The data have a determine/compare ratio close to $2/1$, a graph/formula ratio close to $2/1$, and a positive/negative slope ratio close to $1/1$. Because of the inclusion of all items accompanied by an explanation, about 13% of the kinematics questions in the data set are comparison items. This means that the frequencies provided are predominantly from questions with a determine task, that half of them use a graph and the other half a formula, and that they are almost exclusively from questions with a negative slope.

The most common strategy—and simultaneously the most common error—in general, is identification of the x intercept (I2), with a frequency of 27.1%. A very large difference between the contexts is found: only 9.7% in kinematics and 41.5% in mathematics. The

change of representation also shows interesting results. Constructing an equation (I3) when a graph is provided also shows a stark difference between kinematics (4.1%) and mathematics (20.2%) with an overall frequency of 12.9%. Furthermore, when construction of a graph (I4) is used, it mainly happens in kinematics. Tables are rarely used to determine the y intercept. The most used strategy in kinematics is the determination of the intersection with the vertical axis (I6) of a graph with a frequency of 39.6%, which is in stark contrast with that for mathematics, which is only 11.7%.

The calculation of a specific function value (I7) is also a frequently used method, especially in mathematics with 28.6%, compared to 17.6% in kinematics. A total of 83% of the answers in I7 are correct, meaning students indeed calculate $f(0)$ in the majority of these cases.

2. Slope

The results for the slope questions are shown in Table V. Answers without any explanation are omitted from this analysis. The data set contains 2501 uses of *at least* one strategy across the twelve questions about slope, 1202 in kinematics and 1299 in mathematics. As shown in Fig. 1, all twelve slope questions require an explanation. The data structure is inline with the structure of Fig. 1. This means that the data have a determine/compare ratio close to $2/1$, a graph/formula ratio close to $2/1$, and a positive/negative slope ratio close to $1/1$.

The most used strategy for slope is calculating the ratio of differences (S8) at 31.4% overall and with a substantial difference between the contexts with 41.5% in kinematics and 22.0% in mathematics. The second most used strategy is more qualitative: reasoning with steepness (S6) at 19.0%, which is more often used in mathematics (22.9%) than in kinematics (14.7%). Triangles are drawn on a graph (S7) much more often in mathematics (19.7%) compared to kinematics (5.6%) whereas calculating the ratio of coordinates (S9) is far more frequent in kinematics (19.7%) than

TABLE V. Frequencies (in %) for the slope questions. f is across both contexts, f_K is in the kinematics context and f_M is in the mathematics context. The percentages are relative to 2501, 1202, and 1299, respectively.

Category	f	f_K	f_M
S1: Location in an equation	15.6	1.6	28.5
S2: Identification of the x intercept	2.4	1.7	3.1
S3: Construct an equation	4.7	1.9	7.3
S4: Construct a graph	1.9	2.4	1.5
S5: Construct a table	1.5	2.0	1.1
S6: Reasoning with steepness	19.0	14.7	22.9
S7: Drawing a triangle on a graph	12.9	5.6	19.7
S8: Ratio of differences	31.4	41.5	22.0
S9: Ratio of coordinates	9.7	19.7	0.5
S99: Other	12.6	19.6	6.2

in mathematics (0.5%). Another stark contrast between contexts is the determination of the slope through the location of the coefficient in an equation (S1) which achieves an overall frequency of 15.6% with 28.5% in mathematics and a mere 1.6% in kinematics. The representational transitions are not often used, only the construction of an equation (S3) in mathematics is notable with a frequency of 7.3% compared to 1.9% in kinematics.

VII. DISCUSSION

The results in Secs. V and VI reveal significant main effects and interaction effects, and substantial differences in strategy between the two contexts. In this section, we discuss these results as well as selected examples from student explanations.

A. *y* intercept

The most difficult kinematics questions for this concept are K9 and K11 which both require determining the initial position when a formula is given. Very similar results are found for the mathematics questions M9 and M11, which illustrates the difficulty students have with interpreting an abstract algebraic expression. We assume that experts are more likely to determine the y intercept by identifying the coefficient of the zeroth order term (I1), but to our knowledge there is not yet any study available to support this claim. Our data show that students make very little use of this method and instead attempt a calculation to do so, or determine the x intercept (I2) instead. Furthermore, the change in the order of the terms in an equation between kinematics and mathematics [design choice (iv)] does not have a noticeable effect, though particularly because strategy I1 is infrequently used.

Concerning representational transitions (I3, I4, I5), we find that many students construct an equation (I3) in mathematics but do this far less in kinematics. This increased use of algebraic expressions in mathematics is in line with the conclusions from Acevedo Nistal, Van Dooren, and Verschaffel [18] who studied students' flexibility in choosing a representation to solve linear function problems. They found that students have a very strong preference or feel very strong pressure to use formulas, which increases with the grade they are in. The categorization scheme could be simplified here by grouping I4 and I5 into a single category due to the low frequencies, but category I3 should definitely be kept separate.

The most common strategy—and simultaneously the most common error—in general, was identification of the root, i.e., (I2) confusing the y intercept with the x intercept with far more errors in the mathematics context. As mentioned before, the large number of errors in mathematics can probably be attributed to a low familiarity with the term function value of 0, which was used in the item

formulation. The low familiarity was indicated by the students, though it is part of their official curriculum. This compares well with the GEE results in Fig. 3, which show a significant difference in average accuracy between kinematics and physics for the y intercept. These results can be linked to the conclusions from Davis [1], which also indicated that students have more difficulties with using the official terminology whereas they were more successful with informal terminology such as "the starting point" in physics questions. The use of terminology such as starting point or "initial position" might also have a negative effect though. A small number of students wrote replies similar to "the initial position is always 0 m" in the kinematics questions. These explanations were categorized in the "other" category (I99) and did not occur in the mathematics questions. In many standard kinematics exercises, students are taught to intelligently choose the reference system such that the initial position is in the origin, which simplifies the equations and the calculation. Quotes such as the one above show that some student believe that the point 0 m or the origin in a position-time graph is always the initial position, almost as if that is its name. This implies that students at this early stage of learning kinematics have difficulties understanding the concept of a reference frame. They seem unaware that shifting a reference frame changes the graph and the equation, but still describes the same physics and the same reality. Such difficulties show that this might be an opportune time—or perhaps already too late—to introduce the first notions of relativity.

Category I99 contains another notable error, although also with low frequency: the use of the coordinate notation, e.g., (x, y) or (t, x), which is most often a combination of the x component of the x intercept and the y component of the y intercept. This occurs in both contexts and most often in questions with a graph with a negative slope. This use of coordinate notation relates to the "Cartesian connection," which is the realization that each point on a line represents an ordered pair that satisfies the equation representing the line [1]. Davis [1] states that the y intercept can be used to promote the Cartesian connection since it can be directly identified in the equation—which we categorized as strategy I1 and assume to be the preferred method by experts. Our results show that a minority (less than 10%) use the Cartesian connection to answer a y intercept or initial position question, but do so incorrectly since we explicitly inquire about the function value or the position, which is a single number, not a coordinate.

In general, the most interesting results concerning the y intercept are found in mathematics since here the accuracies are far lower than in kinematics, students' use of strategies shows more variation and use of informal terminology has a bigger impact. Finally, no specific cases are of note for the y intercept questions, the categorization scheme—when not accounting for the representation transitions—can be used very well to classify the most common errors and strategies.

B. Slope

In this section we first discuss the GEE results and the most common strategies and errors. Next, we zoom in on the interval or point confusions by discussing three different cases and compare between kinematics and mathematics. Afterwards, we have a similar discussion about how students treat the sign of the slope in kinematics and mathematics.

1. General discussion

Questions concerning slope are found to be particularly difficult in the kinematics context with a negative slope and the most difficult when presented with a formula. In determine via formula questions about the slope in kinematics, the average accuracies are extremely low compared to their isomorphic version in mathematics. Furthermore, negative slope questions result in acceptable average accuracies in mathematics, but extremely low average accuracies in kinematics. A large factor of influence in these results is the inclusion of the possible minus sign for the one-dimensional velocity as a criterion for a correct answer. The results show that the large majority of students does not include the minus sign in kinematics, i.e., they consider the magnitude and omit the direction. This pattern is not present in the mathematics questions, so when isomorphic equations and graphs are used, students are far more likely to include the minus sign.

Part of the large overall difference between the two contexts can be explained by the more expertlike use of strategies in mathematics by determining the slope from an equation through locating the coefficient in the expression (S1). This strategy achieved a 28.5% frequency in mathematics compared to a mere 1.6% in kinematics. This shows that students do not see the link between the slope of linear functions and of velocity in $x(t)$ expressions, which indicates a weak link between contexts and a compartmentalization of understanding within contexts.

By far the most preferred method in kinematics is calculating the ratio of differences (S8) (mainly in questions with a graph). The method is not always applied correctly though. After calculating the ratio of differences, students often omit the minus sign (when present) from the result. From students' explanations it is clear that this is triggered by the use of the formula $v = \Delta x / \Delta t$ in kinematics and $a = (y_2 - y_1)/(x_2 - x_1)$ in mathematics. Although this is essentially the same formula, there are a few differences in students' use. In mathematics, the formula is usually used correctly. In physics students usually write down the correct formula but often calculate $v = |\Delta x / \Delta t|$. The second most common error in kinematics—also often after writing down the correct formula—is misinterpretation of the Δ and calculation of the ratio x/t of coordinates (x, t) for some point on the graph (S9). This S9 strategy has a frequency of 19.9% in kinematics but only 0.5% in mathematics. In mathematics,

a lot more students resort to drawing a triangle on a graph (S7), which in almost all cases has a set horizontal length of 1 and a vertical length proportional with the slope—usually also including the correct sign in mathematics. This difference between the two contexts for S7 confirms the useful distinction between calculating the ratio of differences (S8) and drawing a triangle on a graph (S7) in the categorization scheme, despite that both essentially calculate a ratio of deltas. The answers in category S8 are correct in 62% of these cases with errors mainly due to omissions of the minus sign or miscalculations. In contrast, the answers in S9—a frequent strategy in kinematics questions—are always incorrect and contain various interesting errors which will be discussed in Sec. VII B 3. This distribution of students' strategies between S7, S8, and S9—which are all predominantly used in questions with graphs—and the frequent use of the expertlike strategy S1 in mathematics are the most striking results for the slope in this data.

Concerning the representation transitions (S3, S4, S5), the results show very low frequencies. Just like with the y intercept, the construction of an equation (S3) is the dominant one and it is more frequently used in mathematics than in kinematics. A straightforward simplification of the categorization scheme is to take S3, S4, and S5 together in the same category called "change in representation," or at least group S4 and S5 into a single category called "change in representation other than equation."

2. Comparison to similar studies

We can compare our results with the categorization into slope or height, interval or point, and iconic representations from Leinhardt *et al.* [13]. First, we find a negligible number of slope or height confusions in both contexts, hence this is not a separate category in our scheme and we categorize these answers into category S99: Other. Second, we do find a substantial number of interval or point confusions—essentially S9 in our scheme—which almost all occur in kinematics questions. Last, we find only a small number of iconic interpretations which are almost always in addition to another strategy and are also categorized in the other category (S99) due to their low occurrence.

We can also compare our results about the ratio of differences and the ratio of coordinates to those from Wemyss and van Kampen [3] and to those from the follow-up study from Bollen *et al.* [7]. The key differences between our study and the others are that we ask to determine the velocity, whereas they ask to determine the *instantaneous speed*, and that our respondents are 9th grade students in the Flemish educational system whereas their students are first-year university students in the Irish, Flemish, or Basque higher educational system. Table VI provides an overview including the relevant data which can be compared, specific details about this are provided in the footnote [40]. In general, we find that the results from our students in Flemish secondary education

TABLE VI. Comparison between our results (Flanders) and those from Wemyss and van Kampen [3] (W & vK) and those from Bollen *et al.* [7] (KU Leuven in Belgium, UPV/EHU in Basque country in Spain, DCU in Ireland). The categories are $\Delta y/\Delta x$ in which in some way a ratio of differences is used (combined S7 + S8); and y/x in which in some way a ratio of coordinates is used (S9). Frequencies f_K in kinematics and f_M in mathematics are expressed in percentages with respect to the total number n of answered questions.

	Flanders		W & vK		KU Leuven		UPV/EHU		DCU	
	f_K	f_M	f_K	f_M	f_K	f_M	f_K	f_M	f_K	f_M
Category	$(n=433)$	$(n=455)$	$(n=403)$	$(n=355)$	$(n=202)$	$(n=202)$	$(n=121)$	$(n=224)$	$(n=480)$	$(n=355)$
$\Delta y/\Delta x$	63	86	22	57	50	81	46	74	18	57
y/x	27	1	56	25	22	5	30	9	66	25

are similarly structured (lowest and highest frequencies, largest differences, ratios of frequencies) as those from the Flemish undergraduate students at KU Leuven, which is that they use the ratio of differences about twice as often as the ratio of coordinates in kinematics, and that they use the ratio of differences far more often than the ratio of coordinates in mathematics. This shows that the effects from the educational system last from secondary education to higher education. Furthermore, all cohorts show large context gaps with kinematics being the most difficult, indicating that students have difficulties linking their understanding of linear function in mathematics to a similar topic in kinematics.

3. Interval or point confusions

In questions with graphs, our data contain many examples of the well-documented interval or point confusion; in questions with an algebraic representation though, this—to our knowledge—has not yet been well studied. In questions with a determine via formula task, different and kinematics-specific errors occur in which students exhibit interval or point confusion often combined with the use of dimensional arguments. In addition, many of the examples in this section illustrate that different contexts often result in different strategies and errors. Three cases are discussed:

Case 1: Numerical interval or point confusion: A value for one of the variables is entered into the equation and the matching value for the other variable is calculated. The ratio x/t is then calculated. This is illustrated in Fig. 8 in which the student chose $x = 2$, calculated the corresponding value for t and then calculated the ratio of both. Note that this student correctly wrote down the formula for velocity but failed to interpret it correctly.

Case 2: Algebraic interval or point confusion: The student seems to look for the ratio of x over t, which by some is literally written as *distance/time*, while actually *position/time* is calculated. The focus here is to find "an x" and "a t" or "a distance (position)" and "a time" because the formula requires them. In contrast with case 1, the first step here is to find algebraic expressions for x and t instead of finding numerical values. This confusion is illustrated in the answer

to K12 in Fig. 9, where the student also omits the deltas from the formula for velocity and tries to find an algebraic expression for x, for which the student uses the equation of motion. After calculating the ratio, the student is stuck and does not know how to continue to end up with a numerical value. In contrast, in the isomorphic question M12 (also shown in Fig. 9), the student manages to solve the question with an (assumed) expertlike strategy by identifying the slope through its location in the equation (S1).

Another illustration of an algebraic interval or point confusion is given by the answer to K10 in Fig. 10, in which the student first makes clear that the equation of motion provides the expression for x, and then manipulates the equation to find an expression for t. Next, the correct formula for v is written, and the previous two expressions are incorrectly entered in that formula like an interval or point confusion. Finally, the student is stuck. Interestingly, in the comparable question M12—in which only the sign of the slope or velocity is different—the student applies the (assumed) expertlike

FIG. 8. Example of case 1 with question K10 in which the equation $x = 4 + 8t$ was provided: a numerical interval or point confusion and additionally the Δ in the formula for v is written but not taken into account.

K12

$$\nu = \frac{x}{t}$$

$$\nu = \frac{3 - 12t}{t}$$

M12

$$f(x) = \underline{-6}x + 12.$$

rico = −6

de rico staat voor de x
in de norm $f(x) = mx + q$

FIG. 9. Example of case 2 with question K12 in which the equation $x = 3{-}12t$ was provided: an algebraic interval or point confusion in which the equation of motion is substituted in the formula for velocity. Additionally, the isomorphic question M12 from the same student is shown, in which strategy S1 is used: the location of −6 is underlined in the question and the explanation translates as "the slope is located before the x in the form $f(x) = mx + q$."

K10

$$\underline{g(x) = 4 + 8t}$$
afstand

$$8t = 4 - x$$
⇓
$$t = \frac{4 - x}{8}$$

$$v = \frac{\Delta x}{\Delta t} \Rightarrow \frac{4 + 8t}{1} \circ \frac{4 - x}{8}$$

$$\left(\frac{4 + 8t}{1}\right) \cdot \left(\frac{8}{4 - x}\right)$$

$$= \frac{32 + 64t}{4 - x} \quad ?$$

M12

$$g(x) = mx + q$$
$$g(x) = -6x + 12$$

de richtingscoëfficiënt is −6

FIG. 10. Example of case 2 with question K10 in which equation $x = 4 + 8t$ was provided: an algebraic interval or point confusion in which the equation of motion is used for x and a manipulated version for t. The comparable question M12 from the same student is solved using strategy S1 by identifying the slope through its location in the equation, made clear by comparing with a standard form $f(x) = mx + q$. The written explanation translates as: "the slope is -6."

K10

$$v = \frac{\Delta x}{\Delta t}$$

$$v = \frac{(4 + 8t) \, m}{(8t) \, s}$$

$$= 4 \, m/s$$

FIG. 11. Example of case 3 with question K10 in which equation $x = 4 + 8t$ was provided: a unit-based interval or point confusion in which the equation of motion and a term in the equation are thought to be expressed in units of meters and units of seconds, respectively.

strategy S1 in which the slope is identified through its location in the equation.

Case 3: Unit-based interval or point confusion: Here, the student tries to find a ratio of something with units in meters over something with units in seconds to achieve meter per second which they know is a proper unit for velocity. Recall that we opted to not include the units in an equation in kinematics, as stated in design choice (iii). Figure 11 shows an example in which the student tries to find an expression in units of meters and an expression in units of seconds and then takes the ratio. For the numerator the equation of motion is chosen since this results in a value expressed in meters. Also, the student incorrectly indicates that the term $8t$ has units of seconds and chooses the full term (coefficient and variable in this particular case) for the denominator. An incorrect calculation then results in 4 m/s.

Figure 12 shows another example of question K10 in which the student identifies the suspected units of the coefficients and takes the ratio to achieve a value with units m/s. The term $8t$ is incorrectly thought to be expressed in units of seconds and the other term is correctly thought to be expressed in units of meters.

These cases and the examples illustrating them, also provide evidence of poor algebraic manipulation skills, as

K10

$$4 \, m \, / \, 8 \, o$$

$$x = 4 + 8t$$
$$\downarrow \quad\quad \downarrow$$
$$m \quad\quad s$$

FIG. 12. Example of case 3 with question K10 in which equation $x = 4 + 8t$ was provided: a unit-based interval or point confusion in which the coefficients from the equation were each linked with a unit and then the appropriate ratio was taken to calculate the velocity.

well as poor understanding of units in algebraic expressions, of the Δ symbol and of the equation of motion. This highlights the weak link between students' understanding in mathematics and in kinematics. The quality of the link between mathematics and kinematics is likely low in case 1 and case 2, since these are basically mathematical approaches without any sign of proper understanding in physics other than a formula learned by heart. Yet at the same time the kinematics context does cause difficulties that do not arise in the isomorphic mathematical questions. In case 3, the students take more of a physics approach since they take the units into account—which is promoted by many teachers—but here they fail to connect the units to the correct parts in the algebraic expression, which highlights their difficulties with algebraic expressions in contexts other than mathematics.

4. The sign of the slope

In questions with a negative slope, and mainly in kinematics questions, students sometimes made statements related to the motion of the cyclist such as: "the cyclist is returning" or "is riding backwards" or "is slowing down." The first two statements indicate that these students include a sense of direction in their interpretation, but some have difficulty to understand and/or express moving with or opposite the direction of the position axis. There was even a respondent who repeatedly wrote "riding backwards or back" thus illustrating doubts about the correct interpretation. The "slowing down" statement illustrates the difficulty students have with interpreting $x(t)$ graphs with constant velocity by confusing them with $v(t)$ or $v(x)$ graphs. Our results show that the sign of the slope in mathematics is no problem for the majority of students, but when confronted with a negative sign for the velocity in kinematics they frequently omit the minus sign. Recall that Dutch speaking students only have the single word snelheid, whereas English speaking students have velocity and speed.

This difference in student strategy between mathematics and kinematics concerning the negative slope is well illustrated with some examples from the test. Figure 13 shows isomorphic questions K8 and M8 solved by the same student, both solved by using a triangle (S7). In kinematics the minus sign is omitted, but is taken into account in mathematics. Also note that in kinematics the students used the intercepts while in mathematics a horizontal step size of $+1$ was used.

Figure 14 also shows two isomorphic questions solved by the same student, but in this case with positive slope. The kinematics question is solved incorrectly and shows an interval or point confusion, while the mathematics question is solved correctly by drawing a triangle (S7) with a horizontal step size of $+1$. In this case M6 was eventually also incorrect because the ratio was inverted.

Figure 14 shows isomorphic questions K8 and M8 solved by the same student who answered the questions
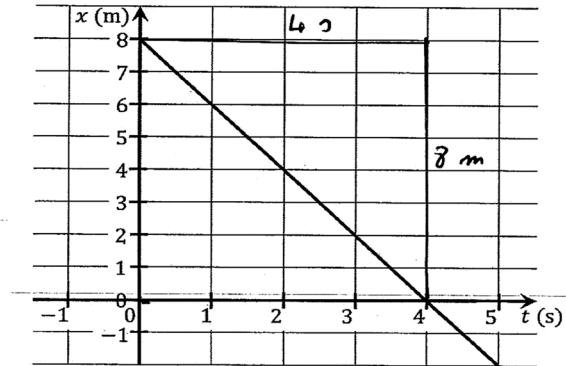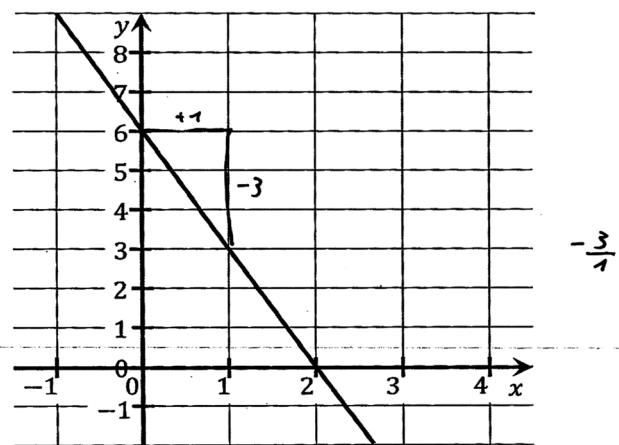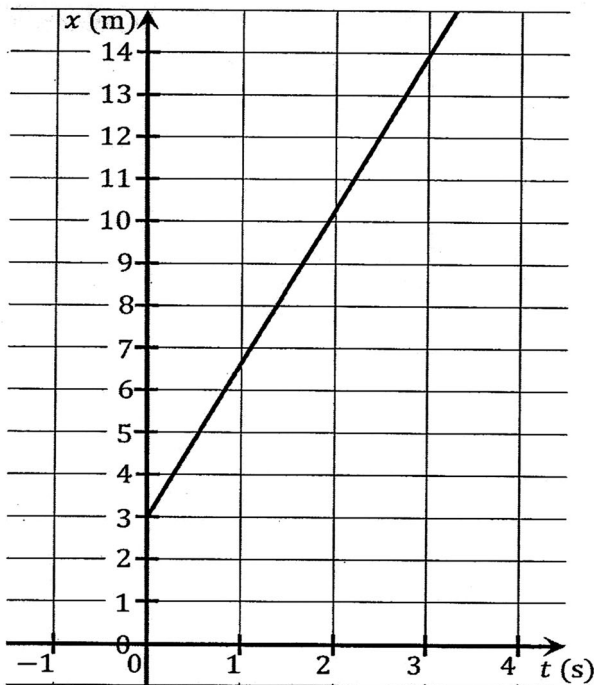

FIG. 13. Example of isomorphic questions K8 and M8 solved by the same student. In both cases strategy S7 is used, but in kinematics the minus sign has not been taken into account in contrast to M8. Translated wording in K8 and M8 are "The speed or velocity of the car is 0.5 meters per second." and "The slope is -3."

in Fig. 15. This student incorrectly solved the kinematics question (interval or point confusion) and omitted the minus sign, but answered the mathematics question by drawing a triangle (S7) with a horizontal step size of $+1$ and includes the minus sign. Again, the mathematics question was also incorrect because this student consistently used the inverted ratio for the slope.

Another question which challenged students' reasoning with negative velocities is K4 in which two negative velocities are compared graphically. As stated before, the average scores were very low for K4, but they were also at the low end for M4. The reason why both were not solved very well is because when asking to compare two negative numbers, e.g., -5 and -2, many students reasoned that -5 was the largest because $5 > 2$, thus not taking into account

K6

3 sec = 14 m $\frac{14 \, m}{3 \, 1} = 4{,}6$

→ 4,6 m\sec = snelheid trein

M6
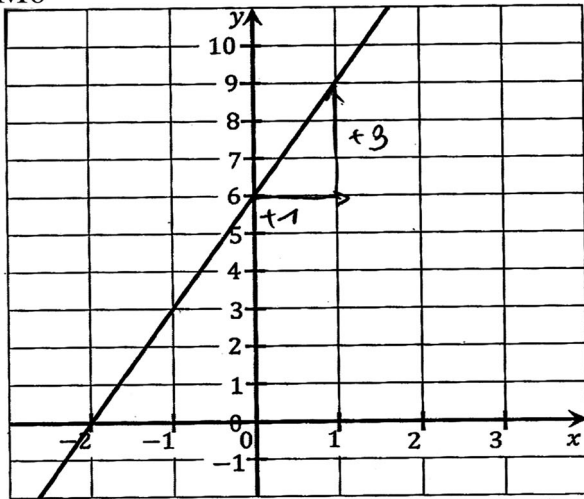
Bepaal de richtingscoëfficiënt. <u>Verklaar</u> je antwoord.
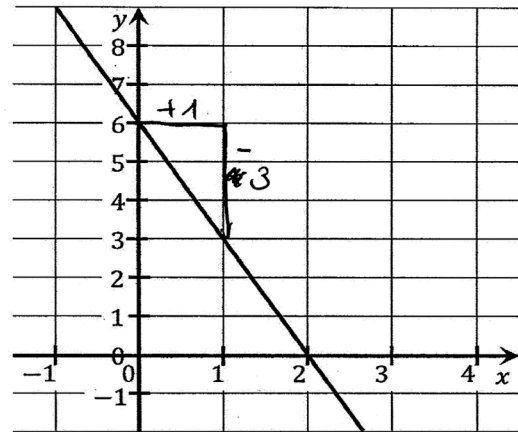
$\frac{1}{3} = 0{,}33\ldots$

FIG. 14. Example of isomorphic questions K6 and M6 (with positive slopes) solved by the same student. In K6, the interval or point confusion occurs resulting in an incorrect answer, but in M6 a triangle is drawn (S7) with a horizontal step size of $+1$ resulting in a correct answer. The text in the K6 translates as "speed or velocity train."

K8
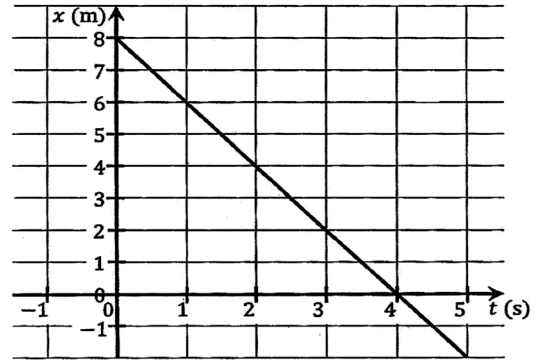
snelheid: $\frac{afstand}{tijd}$ → $\frac{4 \, m}{2 \, sec} = \frac{2 \, m}{1 \, sec}$

→ 2 m\sec = snelheid

M8

$\frac{1}{3} = -0{,}33\ldots = $ richtingscoëfficiënt

FIG. 15. Example of isomorphic questions K8 and M8 (with negative slopes) solved by the same student who answered the questions in Fig. 14. In K8, the interval or point confusion occurs and the minus sign is omitted. In M8 a triangle with horizontal step size of $+1$ is drawn and the minus sign is included. In K8, the text translates as "*velocity: distance/time.*"

the minus sign. This occurred in kinematics and in mathematics. An additional problem for kinematics—as made clear in the discussions above—is that students tend to consider speed instead of velocity, which increased the number of incorrect answers significantly. These two difficulties fully explain the difference between the average accuracies for K4 and M4. No such issues occurred in questions K2 and M2 which are the isomorphic versions with a positive slope.

VIII. IMPLICATIONS FOR TEACHING AND FUTURE RESEARCH

The validation shows that teachers can use the test for groups or individuals [37,41] to assess students'

understanding, to identify difficulties, and to uncover solution strategies and confusions, which can then be used for targeted remediation. The test can also be used as a pretest and post-test for intervention studies similar to how Hill *et al.* [42] used the Force Motion Concept Evaluation [43] and the Representational Fluency Survey [24] to measure learning gains after an intervention. The test can also be modified to suit other research questions in which case some interesting changes can be made, e.g., require an explanation in every question, include negative values for some variables in the graphs, drop some items in favor of items with other sign combinations or other quadrants, expand the test to include other linear graphs from 1D kinematics $[v(t), v(x), a(t), a(x), …]$, include other contexts, etc.

Our comparison with other studies shows substantial differences between different cohorts depending on age group, the educational system and the type of course (algebra- or calculus-based). Because of our fairly large number of respondents, distributed over multiple schools and classes across Flanders, the results are representative for students in a science oriented curriculum in the early stages of learning about kinematics and linear functions. Additional studies with different age groups and curriculum programs would be welcome, as well as longitudinal studies which follow students' progress concerning this topic.

To improve students' understanding and their linking between contexts, we suggest introducing isomorphic quantities such as slope and velocity in parallel. Comparing isomorphic situations from different contexts could be a good approach, e.g., show a graph and the matching algebraic expression in mathematics and in kinematics side by side and annotate them while highlighting the similarities. Additionally, we suggest discussing examples in different reference systems to illustrate the effect on equations and graphs. We also observed difficulties with the Δ symbol, which is why we advise to explicitly write the incremental steps in calculations to clarify its meaning in both contexts, e.g., $v = \Delta x/\Delta t = (x_2 - x_1)/(t_2 - t_1)$. Furthermore, the significant main and interaction effects in our study strongly imply that students could benefit from an increased focus on linking different contexts; therefore, this issue also requires appropriate attention during teacher education.

Because of the isomorphic structure of our test and the direct mapping between, e.g., item M1 and item K1, the test can also be considered a transfer of learning test. Transfer of learning is often defined as the ability to use and apply skills and knowledge in a different context from the one in which they were learned [44–46]. To quantify transfer of learning is not a straightforward task, but Britton *et al.* [47] have made progress by introducing the transfer rating [47], which was subsequently improved by Roberts *et al.* [46] when they introduced the transfer index. Based upon correlating performance between matching questions in

TABLE VII. Allocation of transfer score [46].

	Accuracies			
Item accuracy in context A	1	0	1	0
Item accuracy in context B	1	1	0	0
Transfer score from A to B	2	1	0	0

two contexts, a transfer score can be ascribed to each matching set using the system presented in Table VII. For a test with n pairs of matching items, the transfer index is calculated as follows:

$$\text{Transfer index} = \frac{\sum_{i=1}^{n} \text{transfer score}}{2n} \times 100, \quad (4)$$

resulting in a value between 0 and 100, with higher values indicating higher transfer of learning.

Calculating the transfer index from mathematics to kinematics for our data we find a value of 37.63 ± 17.45. Although it is always possible to calculate this index, a consistent interpretation requires an appropriate research design. Our design was chosen to study how well students in the Flemish education system—which uses mandatory learning goals—perform in two related contexts without controlling for the specific implementation in the classroom, essentially considering the classroom a black box governed by the educational system. This means that our data are not fully suited for interpretation in terms of transfer of learning since we cannot verify one of the important conditions to assess transfer between contexts, which in our study would be that learning took place in mathematics *before* it took place in kinematics. Though it is highly likely to be the case, we did not include this information in our original design. Furthermore, as discussed in Roberts *et al.* [46], to verify the causal relationship of improvements in a context due to transfer of learning from another context, all other variables which can influence the performance should be checked and included, which is not possible with our design. For illustrative purposes, we also calculated the transfer index from kinematics to mathematics which results in 41.58 ± 21.21, which is actually higher than that for mathematics to kinematics. Furthermore, we found a Spearman's rank correlation coefficient ρ of 0.356 between kinematics and mathematics items. These results show that the interpretation of the transfer index is not always straightforward and should be done with appropriate care. With a matching research design and detailed control of classroom activities though, our test can certainly be used in combination with this index to study transfer of learning, which would be an interesting subject for future research. Additionally, it would also be interesting to calculate the transfer index between mathematics and kinematics for, e.g., the items concerning graphs with those concerning formulas separately, which could provide more insight in the effect of representations on transfer.

IX. CONCLUSION

In conclusion, we studied student understanding of the concepts y intercept and slope in linear function problems with graphs and algebraic expressions in isomorphic kinematics and mathematics questions. Test validation resulted in good values for internal consistency, item difficulty indexes, item discrimination indexes and Ferguson's delta. Our results show that students' difficulties are concentrated in physics, algebraic expressions, and questions with negative slope. Results from a GEE analysis of the accuracy show that context (kinematics or mathematics), concept (y intercept or slope), slope sign (positive or negative), and question type (determine via graph, determine via formula or compare via graph) are all highly significant factors, as well as almost all interaction effects (up to three way interactions). The study also resulted in a bottom-up constructed categorization scheme for students' explanations which proved a reliable tool and showed that there are only a few strategies or errors for questions concerning the y intercept, and a more detailed scheme for slope questions. The scheme included representational transitions, but we found that such strategies were not frequently used in general with only one exception, namely, the transition from a graph to an algebraic expression in mathematics questions concerning the y intercept. The main error for the y intercept was the determination of the x intercept instead and students very rarely used the location in an algebraic expression to identify the y intercept. Additionally, we found that students had poor understanding of formal terminology concerning the y intercept since the wording function value of 0 caused some difficulties. More variation was found in students' explanations for the slope questions. For questions with an algebraic expression we found little use of identification of the velocity or the slope through the location in the expression in kinematics but far more in mathematics. Slope or height confusion and iconic interpretations were infrequent in our results, but interval or point confusions were seen often. In graphical questions we confirmed presence of the interval or point confusion in physics, but found almost none in mathematics. In questions with an algebraic expression we also found a high frequency of interval or point confusions in kinematics but almost none in mathematics and discussed three notable cases: numerical interval or point confusion in which the focus lies on finding the numbers for x and t; algebraic interval or point confusion in which the focus lies on finding expressions for x and t; and unit-based interval or point confusion in which the focus lies on finding expressions or values with the requested units for x and t. These were often interwoven with additional errors resulting from poor algebraic manipulation skills, poor understanding of units, poor understanding of the Δ symbol, and poor understanding of the equation of motion. Additionally, isomorphic examples illustrated that there is a weak link between kinematics and physics and that students are unable to successfully transfer their mathematical understanding to kinematics. Finally, we highlighted the sign issue of velocity in kinematics, which for our respondents was a particular issue due to linguistic difficulties. Isomorphic questions again showed that students consider the minus sign in mathematical questions, but ignore the minus sign in kinematics questions, and tend to reason with speed rather than velocity.

[1] J. D. Davis, Real-world contexts, multiple representations, student-invented terminology, and y-intercept, Math. Think. Learn. **9**, 387 (2007).

[2] 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).

[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] M. Planinic, Z. Milin-Sipus, H. Katic, A. Susac, and L. Ivanjek, Comparison of student understanding of line graph slope in physics and mathematics, Int. J. Sci. Math. Educ. **10**, 1393 (2012).

[5] K. Adu-Gyamfi and M. J. Bossé, Processes and reasoning in representations of linear functions, Int. J. Sci. Math. Educ. **12**, 167 (2014).

[6] D. De Bock, W. Van Dooren, and L. Verschaffel, Students' understanding of proportional, inverse proportional, and affine functions: Two studies on the role of external representations, Int. J. Sci. Math. Educ. **13**, 47 (2015).

[7] L. Bollen, M. De Cock, K. Zuza, J. Guisasola, and P. van Kampen, Generalizing a categorization of students' interpretations of linear kinematics graphs, Phys. Rev. Phys. Educ. Res. **12**, 010108 (2016).

[8] S. Ceuppens, J. Deprez, W. Dehaene, and M. De Cock, Design and validation of a test for representational fluency of 9th grade students in physics and mathematics: The case

of linear functions, Phys. Rev. Phys. Educ. Res. **14,** 020105 (2018).

[9] J. Woolnough, How do students learn to apply their mathematical knowledge to interpret graphs in physics?, Res. Sci. Educ. **30,** 259 (2000).

[10] D. E. Trowbridge and L. C. McDermott, Investigation of student understanding of the concept of velocity in one dimension, Am. J. Phys. **48,** 1020 (1980).

[11] D. E. Trowbridge and L. C. McDermott, Investigation of student understanding of the concept of acceleration in one dimension, Am. J. Phys. **49,** 242 (1981).

[12] 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).

[13] G. Leinhardt, O. Zaslavsky, and M. K. Stein, Functions, graphs, and graphing: Tasks, learning, and teaching, Rev. Educ. Res. **60,** 1 (1990).

[14] R. J. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. **62,** 750 (1994).

[15] P. B. Kohl and N. D. Finkelstein, Student representational competence and self-assessment when solving physics problems, Phys. Rev. ST Phys. Educ. Res. **1,** 010104 (2005).

[16] P. B. Kohl and N. D. Finkelstein, Effect of instructional environment on physics students' representational skills, Phys. Rev. ST Phys. Educ. Res. **2,** 010102 (2006).

[17] P. B. Kohl and N. D. Finkelstein, Patterns of multiple representation use by experts and novices during physics problem solving, Phys. Rev. ST Phys. Educ. Res. **4,** 010111 (2008).

[18] A. A. Nistal, W. Van Dooren, and L. Verschaffel, What counts as a flexible representational choice? An evaluation of students' representational choices to solve linear function problems, Instr. Sci. **40,** 999 (2012).

[19] W. M. Christensen and J R. Thompson, Investigating graphical representations of slope and derivative without a physics context, Phys. Rev. ST Phys. Educ. Res. **8,** 023101 (2012).

[20] M. De Cock, Representation use and strategy choice in physics problem solving, Phys. Rev. ST Phys. Educ. Res. **8,** 020117 (2012).

[21] B. Ibrahim and N. S. Rebello, Representational task formats and problem solving strategies in kinematics and work, Phys. Rev. ST Phys. Educ. Res. **8,** 010126 (2012).

[22] A. A. Nistal, W. Van Dooren, and L. Verschaffel, Students reported justifications for their representational choices in linear function problems: An interview study, Educ. Stud. Math. **39,** 104 (2013).

[23] M. T. H. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5,** 121 (1981).

[24] M. Hill, M. D. Sharma, J. O'Byrne, and J. Airey, Developing and evaluating a survey for representational fluency in science, Int. J. Innovation Sci. Math. Educ. (formerly CAL-laborate International) **22,** 22 (2014).

[25] M. Hill and M. D. Sharma, Students representational fluency at university: A cross-sectional measure of how multiple representations are used by physics students using the representational fluency survey, Eurasia J. Math. Sci. Technol. Educ. **11,** 1633 (2015).

[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.15.010101 for the full test translated from Dutch to English. The items are in the order in which they were administered. The first twelve items are in the context of kinematics, the last twelve items are in the context of mathematics. Item numbers match those used in the paper.

[27] D. Hestenes and M. Wells, A mechanics baseline test, Phys. Teach. **30,** 159 (1992).

[28] L. Ivanjek, A. Susac, M. Planinic, A. Andrasevic, and Z. Milin-Sipus, Student reasoning about graphs in different contexts, Phys. Rev. Phys. Educ. Res. **12,** 010106 (2016).

[29] M. Bassok and K. J. Holyoak, Interdomain transfer between isomorphic topics in algebra and physics, J. Exp. Psychol. Learn. Mem. Cogn. **15,** 153 (1989).

[30] E. B. Pollock, J. R. Thompson, and D. B. Mountcastle, Student understanding of the physics and mathematics of process variables in pv diagrams, AIP Conf. Proc. **951,** 168 (2007).

[31] P. Barniol and G. Zavala, Force, velocity, and work: The effects of different contexts on students' understanding of vector concepts using isomorphic problems, Phys. Rev. ST Phys. Educ. Res. **10,** 020115 (2014).

[32] G. Zavala and P. Barniol, A detailed analysis of isomorphic problems: The case of vector problems, in *Proceedings of the Physics Education Research Conference 2013, Portland, OR,* (AIP, New York, 2013), pp. 377–380.

[33] K.-Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika **73,** 13 (1986).

[34] M. Stokes, C. Davis, and G. Koch, *Categorical Data Analysis Using the Sas®System*, 2nd ed. (SAS Institute Inc., Cary, NC, 2000).

[35] C. Sheu, Regression analysis of correlated binary outcomes, Behav. Res. Meth. Instrum. Comput. **32,** 269 (2000).

[36] D. E. Meltzer, Relation between students problem-solving performance and representational format, Am. J. Phys. **73,** 463 (2005).

[37] P. Engelhardt, An introduction to classical test theory as applied to conceptual multiple-choice tests, in *Getting Started in PER*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009).

[38] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. **20,** 37 (1960).

[39] A. J. Viera, J. M. Garrett *et al.*, Understanding interobserver agreement: The kappa statistic, Fam. Med. **37,** 360 (2005).

[40] The questions selected in our test are all those with the "determine via graph" question type in which the concept is the slope or velocity. This includes all kinematics and mathematics questions and both positive and negative slope. For detailed comparison between the data from the other two studies we refer to the study from Bollen *et al.* [7]. Note that KU Leuven and UPV/EHU data is from undergraduate students in a calculus-based course and that from W & vK and DCU are from students in an algebra-based course at DCU. The selected data from the two referenced studies is from Table II and Table III in each of them. This includes all questions with distance-time graphs

(kinematics) in which the respondents are asked to determine the *speed* at a specific *instant*, and all $y(x)$ graphs (mathematics, which they call context-free) in which the respondents are asked to determine the slope at a specific *point*. In our data, questions without an explanation are not included. In the referenced studies though, there is a category "No answer" in the mathematics data. For a fair comparison, we subtract the cases in no answer from the total and recalculate the frequency. In their kinematics questions there is a category "Other or no answer" or "No response or incoherent or other," which complicates the comparison. In the study from Wemyss and van Kampen we see that in the mathematics table the total prevalences in "Other" and "No answer" are very similar. In line with this we subtract half of the other or no answer prevalence from the total n in the kinematics table. We did the same with the data from the Bollen *et al.* This seems appropriate for the DCU data since the numbers in other and no answer are very similar, but less so for KU Leuven and UPV/EHU. For these last two universities, we calculate that the maximum deviation is only about 3%, which would occur if, e.g., "No response or incoherent or other" would actually all be no response, which is unlikely. There are two useful categories to compare. The first are the ones in which some form of a ratio of differences is used—which are S7 and S8 in our scheme—and their category $\Delta y/\Delta x$ and $\Delta x/\Delta t$. The

second one is the one in which some form of a ratio of coordinates is used, meaning S9 in our scheme and y/x and x/t in theirs.

[41] R. L. Doran, *Basic Measurement and Evaluation of Science Instruction* (National Science Teachers Association, Washington, DC, 1980).

[42] M. Hill, M. D. Sharma, and H. Johnston, How online learning modules can improve the representational fluency and conceptual understanding of university physics students, Eur. J. Phys. **36**, 045019 (2015).

[43] R. K. Thornton and D. R. Sokoloff, Assessing student learning of newtons laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66**, 338 (1998).

[44] D. N. Perkins and G. Salomon, *Transfer of Learning* (Pergamon Press, Oxford, England, 1992) pp. 6452–6457.

[45] J. Mestre, Transfer of learning: Issues and research agenda (2002).

[46] A. L. Roberts, M. D. Sharma, S. Britton, and P. B. New, An index to measure the ability of first year science students to transfer mathematics, Int. J. Math. Educ. Sci. Technol. **38**, 429 (2007).

[47] S. Britton, P. B. New, M. D. Sharma, and D. Yardley, A case study of the transfer of mathematics skills by university students, Int. J. Math. Educ. Sci. Technol. **36**, 1 (2005).