Translating between graphs and equations

Van den Eynde, Sofie; Kampen, Paul van; Van Dooren, Wim ; De Cock, Mieke

Published in:
Physical Review Physics Education Research

DOI:
10.1103/PhysRevPhysEducRes.15.020113

Publication date:
2019

Translating between graphs and equations: The influence of context, direction of translation, and function type

Sofie Van den Eynde,[1,*] Paul van Kampen,[2,†] Wim Van Dooren,[3,‡] and Mieke De Cock[1,§]

[1]*Department of Physics and Astronomy, KU Leuven, Celestijnenlaan 200c, 3001 Leuven, Belgium*
[2]*Centre for the Advancement of STEM Teaching and Learning & School of Physical Sciences, Dublin City University, Glasnevin, Dublin 9, Ireland*
[3]*Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, 3000 Leuven, Belgium*

(Received 30 November 2018; published 10 July 2019)

We report on a study investigating the influence of context, direction of translation, and function type on undergraduate students' ability to translate between graphical and symbolic representations of mathematical relations. Students from an algebra-based and a calculus-based physics course were asked to solve multiple-choice items in which they had to link a graph to an equation or vice versa and explain their answer. The first part of the study focuses on the accuracy of the chosen alternative. Using a generalized estimating of equations (GEE) analysis we find that mathematics items are solved better than physics or kinematics items; that items starting from a graph are solved better than those starting from an equation; and that items on inversely proportional functions are the hardest for students. Quantitatively we see big differences in the number of correct answers between the algebra-based and calculus-based cohorts, but qualitatively the effects of the item variables on the accuracy are the same for both groups. The second part of the study focuses on students' argumentations for the chosen alternative. We observe that students use their physical knowledge 2 to even 3 times as much in kinematics items than in other physics items. When there is an effect of context on argument use, we observe that the use of an argument in a mathematics context differs significantly from the use of that argument in the physics and/or kinematics context. With regard to function type, students explain their choice of answer in items on inversely proportional functions with different arguments and fail more often to answer correctly. In general, the data also show that students from the calculus-based course use more mathematical arguments and score better on the items.

DOI: 10.1103/PhysRevPhysEducRes.15.020113

I. INTRODUCTION

Physics concepts are expressed using external representations (from now on called representations): equations, graphs, tables, diagrams, etc. Since different types of representation may stress different aspects of a physics concept; interpreting and using different representations and coordinating multiple representations is highly valued in physics, both as a tool for understanding concepts and as a means to facilitate problem solving [1]. To benefit from using different representations, students need both representational fluency (the ability to interpret or construct representations [2], and to translate and switch between representations accurately and quickly [1,3]) and representational flexibility (making appropriate representational choices in a given problem solving or learning situation [1,4]).

In this article, we consider students' representational fluency in translating between graphical representations (graphs) and symbolic representations (equations) of different relations between variables in mathematics, kinematics, and other physics contexts. Based on the literature, we expect that context is a first important factor in the way students solve problems. Research on the relation between mathematics and physics has received a lot of attention in physics education research (PER) [5]. Several studies have investigated how students solve problems in mathematics and physics [6–11]. In these studies, students mostly encounter fewer difficulties with mathematics items than with physics items. Students experience a variety of problems when applying mathematical ideas, techniques or structures in a physics context. This is in part due to the different way symbols are used to make meaning and the formal syntax being different in physics and mathematics [12]. As a result, interpreting these symbols is a common source of difficulties for (physics) students. In physics, each

*sofie.vandeneynde@kuleuven.be
†paul.van.Kampen@dcu.ie
‡wim.vandooren@kuleuven.be
§mieke.decock@kuleuven.be

symbol stands for a physical quantity with its own meaning. So, even if students have learned the relevant mathematical tools in their math courses, they still need to learn a component of physics expertise not present in math class: tying those formal mathematical tools to physical meaning. To succeed in physics, students need not just to be fluent with mathematical processing in the context of physics, but also with the mathematical modeling of physical systems, blending physical meaning with mathematical structures, and interpreting and evaluating results [12].

There is well-documented evidence of student difficulties with graphing *per se* [13–16]. Student difficulties with graphs in different contexts have also been studied intensively, especially in the area of kinematics [1,7,17–25]. In a study on the impact of context on student understanding of slope and area under the graph, Planinic *et al.* [17] included three contexts: mathematics, physics (kinematics), and contexts other than physics (e.g., economics). Mathematics seemed to be the easiest context for students, but no significant difference was found between physics and the other contexts. Wemyss and van Kampen [7] used isomorphic questions to investigate the difficulties students in an algebra-based physics course encountered with determining from a graphical representation (a linear graph that did not pass through the origin): (i) the slope of a y, x graph, (ii) the speed and direction of a ball, and (iii) the rate at which the water level in a swimming pool increased or decreased. Most of their students correctly determined the slope of a y, x graph; less than half correctly determined the speed of a ball, but more students correctly determined the rate at which the water level in a swimming pool changed. Their results imply that context is an important factor in problem solving. Bollen *et al.* [25] generalized this study to other educational contexts, and found that students in a calculus-based physics course also did better on the context-free y, x graphs, but did not answer the kinematics and water level questions differently.

Research on translating between and choosing from different representations including graphical representations [26–28] is more scarce. De Bock *et al.* [29] considered equations, graphs, and tables in their study of first-year undergraduate students' understanding of directly and inversely proportional and affine functions, which are three commonly used representations in both physics and mathematics. They found that translating from equation to graph and vice versa is the most difficult for students. Items in which a table is involved are the easiest ones for the students.

A third factor that we expect to have an influence on representational fluency, besides the direction of translation and context, is function type. When De Bock *et al.* [29] investigated students' understanding of proportional, affine (linear functions with a nonzero y intercept), and inversely proportional functions, they found that many students assumed linearity in situations where the relation is not linear, or a wrong linear relation (e.g., proportional instead of affine). Regarding functions, for instance, the straight line graph prototype proved to be very appealing for many students, even if the relation is not linear. De Bock *et al.* found that items on directly proportional and increasing affine functions are solved best by the students.

In this paper, we further explore representational fluency. In particular, we study to what extent students can link symbolic and graphical representations of different function types and whether context, direction of translation and function type have an effect on that. As in the research of De Bock *et al.* [29], we consider directly proportional, inversely proportional, increasing affine, and decreasing affine functions. We choose these four functions because these are the most common functions used in secondary school and introductory mathematics and physics courses. We consider three contexts: mathematics, kinematics, and physics other than kinematics (which will be called physics in the remainder of the article). In this study, mathematics items can be considered as items without context. The distinction between kinematics and physics may give interesting results because of the unknown influence of prior learning, which is typically more extensive for kinematics. We look at the fluency in translating both from equation to graph and from graph to equation, and we consider two qualitatively different cohorts of students: an algebra-based and a calculus-based class. We also investigate whether student problem solving strategies depend on those three item variables (context, direction of translation, and function type).

We therefore address the following research questions:

1. How accurate are students in translating between graphical and symbolic representations of selected functions?
   (a) What is the influence of context [physics, kinematics, or no context (mathematics)]?
   (b) What is the influence of the direction of translation (from symbolic to graphical representation or vice versa)?
   (c) What is the influence of function type (proportional, inversely proportional or affine)?
2. How are the arguments students use when explaining their response on the item affected by context, direction of translation, or function type?
3. Are there qualitative differences between the results of students attending an algebra- and a calculus-based course?

II. METHOD

To answer the research questions, a paper-and-pencil test was developed and administered to first-year university students. In this section, we describe the participating students, the test design, and the methods of analysis.

A. Participants and educational context

Data were collected at two institutions (DCU and KU Leuven). The test was first taken by 306 first-year undergraduate students at DCU in Dublin, Ireland. These students studied science, but did not major in mathematics or physics. There are roughly equal numbers of male and female students who are predominately white. The demographics are broadly similar to that of the local population. They all participated in the same algebra-based introductory physics course. The course provided an introduction to the physics topics of mechanics, fluids, heat, and temperature directed towards general science students. The course consisted of mostly interactive lectures in which students were asked to think about concepts first alone and then in pairs. During the same semester the students also took a mathematics course, Mathematics I (calculus). This year-long module reviewed foundation mathematics and developed the computational skills and techniques of differential and integral calculus. At the time of the test, the students had learned about functions and limits, but not yet about differentiation or integration.

The same test was later administered at KU Leuven in Flanders (Belgium) in the second semester. The questions were also in English, but students were allowed to formulate their answers in Dutch. The students were used to reading physics questions in English because their physics textbook is also in English. These students were therefore familiar with the specific terminology mostly in English. Two-hundred thirty-six first-year undergraduate students in pharmaceutical sciences participated. The student population is mostly female and predominately white, which is representative for the undergraduate population in this area of study in Belgium. They were all enrolled in the same introductory calculus-based physics course, Physics with Elements of Mathematics II. The topics treated in this course are electrostatics, electrical currents, magnetic fields, time dependent fields, and electromagnetic waves. The students do not have a separate mathematics course in their undergraduate curriculum, but they had a physics course preceding this one: Physics with Elements of Mathematics I. The subjects treated in this course are mechanics (including kinematics), oscillations, waves, and sound, Bohr's model and Schrödinger's description of the (quantum) atom, thermal physics, hydrodynamics, and geometrical optics. The course in Leuven consisted of traditional lectures using concept tests and peer instruction, lab sessions, and exercise sessions.

Both sets of students completed the test under exam conditions: they could not communicate with each other and could not consult any textbook or lecture notes. Both institutions are inclusive institutions in terms of their undergraduate admissions practices.

While there were many differences in cultural and educational backgrounds between the two sets of students, we chose to refer to the sets of students from DCU as "algebra-based" and those from KU Leuven as "calculus-based." We hypothesize that these results are probably generalizable to different cohorts, and labels such as "DCU" and "KU Leuven" are probably meaningless to the average reader.

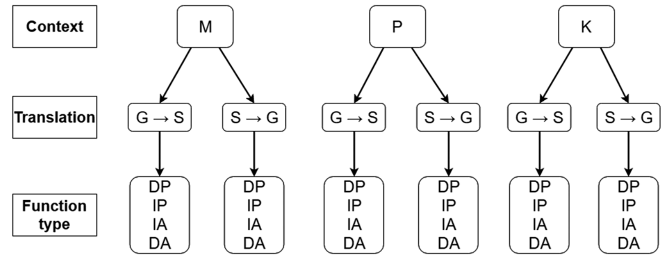


FIG. 1. Schematic overview of the structure of the items.

B. Test design

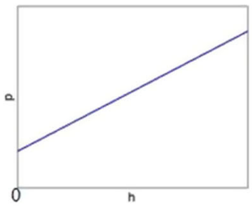
Since we investigate the influence of three contexts, two translations, and four function types, we designed 24 different test items. Figure 1 represents diagrammatically the following branching of test items:

- Context: mathematics (M), physics (P) and kinematics (K);
- Translation: graphical to symbolic ($G \rightarrow S$) and symbolic to graphical ($S \rightarrow G$);
- Function type: directly proportional (DP), inversely proportional (IP), increasing affine (IA), and decreasing affine (DA) functions.

First the four function types were selected, which mirrors those investigated by De Bock *et al.* [29]. These functions are selected because these are the four most common function types students know from secondary school. We then searched for an appropriate physics context for each function type, starting from the physics curricula in Ireland and Belgium in secondary school and higher education as a source of inspiration for possible topics. From these we chose four topics that were familiar for both cohorts: Ohm's law (inversely proportional, resistance R versus current I), Hooke's law (directly proportional, force F versus extension x), hydrostatic pressure (increasing affine with positive intercept with the vertical axis, pressure p versus depth h), and thermal expansion (decreasing affine, the length L versus the temperature difference ΔT). These four physics contexts were selected because both student cohorts had some prior knowledge of them.

At DCU, most students (between 75% and 80%) took physics as part of a general science course at lower secondary level, took no physics at upper secondary level, and had completed 6 weeks of an algebra-based introductory physics course that did not cover any of these four topics. These students would have encountered, at lower secondary level, the variation of pressure in a liquid with depth and thermal expansion only qualitatively, but would have seen and manipulated both symbolic and graphical representations of Ohm's law and Hooke's law. At KU Leuven, all students took physics at the lower and upper

4a. In a container filled with water the hydrostatic pressure p was measured at different depths h. The relationship between hydrostatic pressure and depth is shown in the graph below.



Which equation gives this relationship between hydrostatic pressure and depth, given that p_0, ρ and g are positive constants? Circle a letter a-h.

a) $p = \rho g h$
b) $p = \rho g h + p_0$
c) $p = \rho g h - p_0$
d) $p = \rho g h^2$
e) $p = \rho g$
f) $p = -\rho g h + p_0$
g) $p = \frac{\rho g}{h}$
h) $p = e^{-\rho g h}$

Explain your answer.

FIG. 2. Example of a test item about an increasing affine function in a physics context, translating from a graphical to a symbolic representation.

secondary level, some at a more basic level (25%) and some at a more advanced level (75%). They also completed a calculus-based introductory course that covered some of the topics. All students have encountered Hooke's law, Ohm's Law, hydrostatic pressure, and thermal expansion at some point in their education, definitely in a symbolic representation, and some of them also in a graphical representation.

However, at both institutions there was little to no explicit attention to translating between those different representations.

Each test item consists of a description of the physical situation and the explicit relation between two variables describing that situation. In test items of the type $G \rightarrow S$, the relation is given as a graph, and students choose the corresponding equation from a list of eight options, which are: DP, IA with positive vertical intercept, IA with negative vertical intercept, quadratic, constant, DA, IP, and exponentially decreasing. An example is given in Fig. 2. The same eight options were given in test items of the type $S \rightarrow G$, in which the relation is given as an equation and students have to select one from eight graphs (see Fig. 3). Moreover, students are asked to explain their choice, which allows us to investigate differences in problem solving strategies as there is evidence in the literature that different representations trigger different solution strategies [1,30,31]. We also investigated to what extent context influences the arguments students make. In the research of Wemyss and van Kampen [7] and the generalization of Bollen *et al.* [25], different trends were seen among students in algebra-based and calculus-based classes. Sometimes context had an effect on which strategy is used, and sometimes not.

Next, we developed an isomorphic mathematics and kinematics item for every physics item with similar syntax. All items can be found in the Supplemental Material [32]. Note that the axes had no numerical scale (other than the origin) to ensure students had to tackle the task at a somewhat abstract level. This makes the tasks probably harder than those given by De Bock *et al.* [29], who found that tabulated data posed the least difficulty to their students, and concluded that the use of specific numerical values helps students to reason about a given mathematical situation.

The test items are very similar in phrasing and content. We balanced our desire to obtain as extensive a data set as possible with the need to limit the number of items each student answers to avoid that they see that all questions are inherently the same. We asked every student to solve only four out of twenty-four items, all having the same context and the same translation but having four different function types. Every set of four items exists of two versions with different order, resulting in twelve different tests of four items.

C. Analysis procedure

We analyzed student answers from two different perspectives: the accuracy of the multiple choice answers on the one hand and the explanation on the other hand.

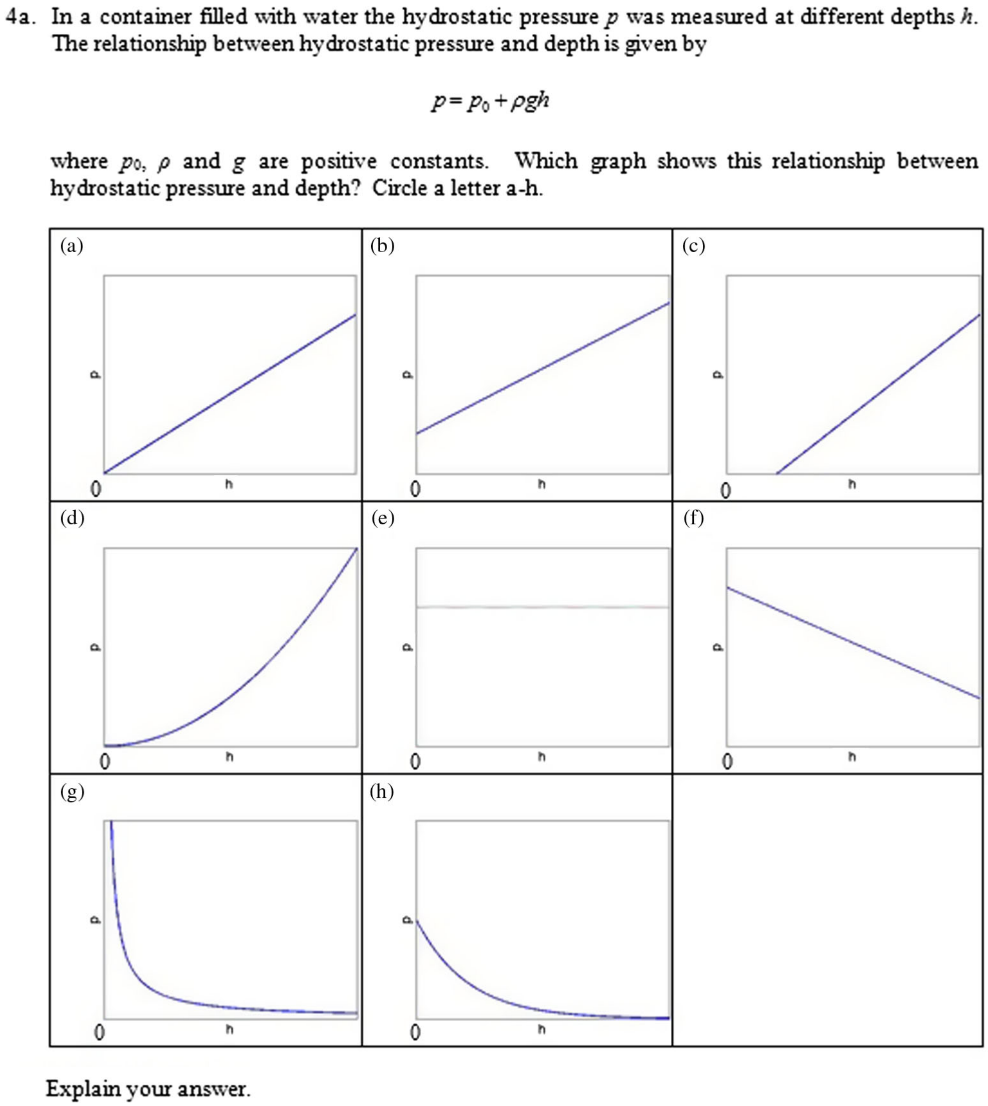


FIG. 3. Example of a test item about an increasing affine function in a physics context, translating from a symbolic to a graphical representation.

With respect to accuracy, student answers are coded dichotomously, i.e., as correct or incorrect, based on their answer on the multiple-choice part of the item. Reasoning is not yet taken into account in this step, so it is possible that students choose the correct alternative but give an incorrect explanation, or vice versa.

The first part of our analysis is quantitative and concerns the significance of each of the three predictors (context, translation, and function type), and possible interactions between them. While the focus of this paper is on the results, the quantitative methods we used to obtain them are perhaps not familiar to all readers. Briefly, we computed pairwise comparisons for all-level combinations of the specified factors (e.g., mathematics, physics, and kinematics in the case of the context parameter). Since the dependent variable (a correct or incorrect response) is not continuous but dichotomous, we cannot use ANOVA. Instead, the responses were statistically analyzed by means of the generalized estimating of equations (GEE) [33] approach within SPSS. We have used a logistic regression in a GEE analysis with three predictors: context ($M/P/K$), translation ($G \to S/S \to G$) and function type (DP/IP/IA/DA). The test statistic of choice for a logistic regression is the so-called Wald chi-squared parameter. To keep the paper focused, we have opted to only quote adjusted significance levels but not odds ratios or risk ratios. In the cases when we perform hypothesis tests with interaction effects, the overall significance level is adjusted for multiple comparisons using a Bonferonni correction.

As we are not only interested in the accuracy of the answers but also in the reasoning students used to solve the problems, their explanations were studied in detail. We used open coding techniques to develop a generative categorization scheme bottom up from the data [34].

Kappa	Agreement
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21– 0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

FIG. 4. An interpretation of Cohen's kappa [35].

We performed interrater reliability tests using percentage of agreement and Cohen's kappa [35]. An interpretation of Cohen's kappa is given in Fig. 4. This way, we checked if the categories are well described and objectively formulated so that an independent rater categorizes the answers in the same way. The construction of the code book is further elaborated upon in Sec. IV A and interrater reliability is discussed in Sec. IV B. Using the categorized student answers, we then once again used the GEE procedure to analyze the effects of the item variables on the probability that certain explanations are given by the students. In this case, the dependent variable was the occurrence of a specific argument type (dichotomous: either used or not used in the explanation) and the predictor variables were again context, direction of translation, and function type.

III. RESULTS: ACCURACY

In the following two subsections, we report on the results of the two cohorts of students described earlier. Both data sets were analyzed using the same method but are kept separately because the educational context of both student populations is different.

Table I shows the percentage of correct answers for the different contexts in the different translations. Table II gives the percentage of correct answers for the different function types in different contexts, while Table III gives the percentage of correct answers for the different function types in the different translations.

A. Main effects

Table IV shows the p values for the main effects of the item variables on the accuracy of student answers.

TABLE I. Percentage of correct answers per context and translation. N is the number of student answers per item type (given in brackets).

		$G \to S$	$S \to G$
Algebra	M	67 ($N = 212$)	59 ($N = 212$)
	P	53 ($N = 208$)	52 ($N = 208$)
	K	57 ($N = 184$)	40 ($N = 200$)
Calculus	M	91 ($N = 148$)	89 ($N = 160$)
	P	74 ($N = 160$)	71 ($N = 156$)
	K	79 ($N = 160$)	74 ($N = 152$)

TABLE II. Percentage of correct answers per context and function type. The last column shows N, the number of student answers per item type.

		DP	IP	IA	DA	N
Algebra	M	63	53	65	72	106
	P	67	41	56	46	104
	K	51	40	56	44	96
Calculus	M	94	73	96	97	77
	P	86	53	83	68	81
	K	87	47	88	83	78

TABLE III. Percentage of correct answers per translation and function type. The last column shows N, the number of student answers per item type.

		DP	IP	IA	DA	N
Algebra	$G \to S$	63	50	68	54	151
	$S \to G$	59	39	50	54	155
Calculus	$G \to S$	90	51	96	87	119
	$S \to G$	88	64	82	79	117

We considered $p < 0.05$ as significant. The logistic regression analysis revealed a main effect of all three explanatory variables (context, direction of translation and function type) for both sets of students. Combining these results with the percentages of accuracy given in Tables I, II, and III, we observed quantitative differences between the two sets (students in the calculus-based cohort more frequently give correct answers), but qualitatively the effects of the item variables were the same:

- Students' accuracy in translating between different representations depended on the context in which the question is asked. Pairwise comparisons showed that mathematics items were solved significantly better than physics and kinematics items, while the accuracy of the latter two did not differ significantly.
- Items requiring translation from graphical to symbolic representation were solved significantly better than items from symbolic to graphical.
- The percentage of correct answers was significantly lower for inversely proportional functions than for the

TABLE IV. p values and Wald χ^2 for the main effects of the item variables on the accuracy of student answers. Significant effects are given in bold.

		Context	Translation	Function
Algebra	p	**0.002**	**0.017**	**0.000**
	Wald χ^2	12.369	5.728	21.404
Calculus	p	**0.000**	**0.047**	**0.000**
	Wald χ^2	16.710	3.952	95.663

TABLE V. p values and Wald χ^2 for the interaction effects between the item variables on the accuracy of student answers. Significant effects are given in bold.

		Context x translation	Context x function	Translation x function
Algebra	p	0.114	**0.037**	**0.022**
	Wald χ^2	4.336	13.393	9.607
Calculus	p	0.886	**0.024**	**0.001**
	Wald χ^2	0.243	14.586	16.588

three other functions, for which accuracies did not differ significantly.

B. Interaction effects

We also investigated the interaction effects between the variables. Simply put, if there is an interaction, the effect of a certain variable is different for different values of another variable. In Table V, the p values for the interaction effects between the item variables on the accuracy of student answers are shown. For both student cohorts, we observed significant interaction effects between context and function type and between direction of translation and function type and no interaction between context and translation.

Combining these results with the percentages of accuracy given in Tables I, II, and III, we again observed quantitative differences between both sets of students, but this time there were also some qualitative differences. The same interactions are significant, but performing pairwise comparisons showed that there are some minor differences in those interactions.

First, we looked at the interaction between context and function.

- For the students from the algebra-based course, items on decreasing affine functions were solved better in mathematics than in physics or kinematics, directly proportional functions were solved better in the physics context than in the kinematics context, but there was no difference between physics and mathematics. For both of the other functions context made no difference.
- For the students from the calculus-based course, items on decreasing affine functions were solved best in mathematics, second best in kinematics and worst in the physics context. Items on inversely proportional functions were solved better in mathematics than in physics or kinematics. Items on increasing affine functions were solved better in mathematics than in the physics context, but there was no difference with the kinematics context. For the items on a directly proportional function, context made no difference.

Second, we did a pairwise comparison for the interaction between function and direction of translation.

- For the students from the algebra-based course, items on inversely proportional and increasing affine functions were solved better in a $G \rightarrow S$ format than in a $S \rightarrow G$ one. For both other functions, direction of translation made no difference.
- For the students from the calculus-based course, items on increasing affine functions were solved better in a $G \rightarrow S$ format than in a $S \rightarrow G$ one. For the three other functions, direction of translation made no difference.

IV. RESULTS: ARGUMENTS

In addition to the accuracy of students' answers, we also looked at the arguments that students used to justify their choices. We developed a categorization scheme bottom up from the data that allowed us to investigate the effect of context, direction of translation, and function on the use of arguments.

A. Code book

Students' argumentations were very fragmented and so categorizing them into all-encompassing categories was not possible. In fact, most students' explanations consisted of multiple parts. Therefore we did not take their whole explanation as the unit of analysis but the separated arguments they use to build it. Based on the data, different types of arguments were distinguished. The arguments that emerged from students' explanations are categorized in three broad groups: graphical, symbolic, and other arguments. We give an overview of the different arguments in those groups and their meaning. Student answers were coded by marking per argument type if it is used (coded 1) in the answer or not (coded 0). It is important to note that it is possible that a certain argument is used incorrectly in the explanation.

1. Graphical arguments

G1.—These are all arguments in which students refer to intersections of the graph with the horizontal or vertical axes or to asymptotes. This type of reasoning can start from the graphical representation, an example is "The graph intersects the y axis in the positive region so…" But it can also start from the symbolic representation, by substituting zero into an equation to determine whether the intersection point is above or below the horizontal axis. An example is "If $F = 0$ then x must also be equal to zero so the graph must go through the origin."

G2.—All the arguments that refer to the shape of the graph. Examples are "it's a curve," "it's (not) a straight line,", "the graph is exponential," and "graph looks like it's a quotient."

G3.—When students mention that the graph is increasing or decreasing or horizontal, but in a graphical way (e.g., not calculating the slope or referring to the symbolical form of the slope). Examples of answers are "The line is not horizontal,", "The graph is moving up so…," and "The graph is decreasing so…"

2. Symbolical arguments

S1.—All arguments related to slope or rate of change. Examples of answers are "There is a constant increase so…" and "a is positive so the graph should be increasing."

S2.—Arguments in which the relation or link between the variables is named, or when the relation is given in symbolic form when asked to translate from G to S. Examples: "they are directly proportional,", "this is a linear equation,", or "first order, so linear."

3. Other types of arguments

D.—Arguments that refer to the (co-)variation [36] of the variables. They often take on the shape of "If … increases, then …. increases/decreases."

I.—Students fill in numbers in the equation or assign values to the axes of the graph.

P.—All arguments that explicitly show the use of physical knowledge. Examples are "If you add more water, the pressure will increase" and "As the bullet is fired it slows down because of gravity."

Q.—Use of the term "positive" or "negative" graph where increasing or decreasing is intended; or using "constant graph" for "graph with constant slope."

4. Examples of coding

Student A answered the item about a directly proportional function in the physics context starting from a symbolic representation with the following explanation:

> *"As **k is a constant**, $F \sim x$, and therefore they are **directly proportional** to each other and so a **positive straight line** is present where **k is the slope**. **As F increases, x increases**."*

We have bolded the passages that made us code this excerpt as $G2, S1, S2$, D and Q.

Student B answered the item about an inversely proportional function in the kinematics context starting from a graphical representation with the following explanation:

> *"**Velocity and time are inversely proportional**, because **when time increases the velocity decreases**. So*
>
> $$\rightarrow v \sim \frac{\textbf{constant}}{\textbf{time}}$$
>
> ***x is a constant** because **all people ran the same distance x**."*

We coded this excerpt as $D, S2$, and P.

Student C answered the item about an increasing affine function in the physics context starting from a graphical representation with the following explanation:

> *"The graph **does not start in the origin, but at a certain height on the y-axis**. So there should be a p_0 where the equation should be added to. The **graph is increasing**, so **the slope is positive**. By eliminating all equations that don't satisfy these conditions, the only possible answer is b."*

We coded this excerpt as $G1, G3$, and *S1*.

B. Interrater agreement

The generative coding scheme was developed bottom up from the data using open coding techniques [34]. Similar methodology in the coding process was used by Ceuppens *et al.* [37]. The tests were first administered in the algebra-based course. The data consist of 1224 student answers. The first author proposed a set of categories. Two authors independently analyzed a subset of the data using this code book (251 of 1224 student answers). We established that there were no missing or ambiguous categories. Then, percentage of agreement and Cohen's kappa were calculated for every argument (see Table VI). We use a minimum value of 80% for the percentage of agreement, and a minimum value of 0.60 for Cohen's kappa to decide that a code is reliable. The cases in which the opinions differed were discussed among the authors until agreement and the code book was revised and updated.

For argument *G3* the kappa is low, but the percentages of agreement is high. A limitation of kappa is that it is affected by the prevalence of the finding under consideration [35]. For rare findings (argument *G3* was only used in 14% of the answers), low values of kappa may not necessarily reflect low rates of overall agreement. Therefore it is important to

TABLE VI. Percentage of agreement between two raters (displayed as % in the table) and Cohen's kappa (displayed as K in the table) for every type of argument, for both the algebra-based and the calculus-based cohorts.

	Algebra		Calculus	
	%	K	%	K
G1	83	0.637	92	0.816
G2	93	0.813	88	0.748
G3	92	0.423	93	0.843
S1	90	0.739	86	0.657
S2	87	0.672	90	0.778
D	84	0.677	85	0.642
I	95	0.621	100	1
P	92	0.781	92	0.740
Q	96	0.723	not applicable	

take into account the combination of both percentage of agreement and Cohen's kappa and the prevalence of arguments. With this revised and updated code book we coded all 1224 student answers of the algebra cohort. An example of a disagreement in coding between authors is given at the end of this section.

When the test was administered in the calculus-based course, we started from the most recently updated code book used for the algebra-based cohort. Again, two authors independently coded a subset of the data (200 out of 944 responses). Then, percentage of agreement and Cohen's kappa were calculated. The cases in which the opinions differed were discussed among the authors until agreement and the code book was revised and updated. The overall categories appeared adaptable to the responses of the calculus-based cohort. However, the code book needed some small additions for answers specific for the calculus-based cohort. Often these were based on the difference in language. The Dutch speaking students sometimes used different jargon, and there were no examples for these cases in the code book. Therefore, extra examples and refinements were added.

In the first round of coding, argument *S2* was the only argument with low kappa and percentage of agreement. The authors discussed the differences in coding, refined the description with some examples (again, typically originating from the different jargon in Dutch). We then selected a new subset of 200 answers to code. These results meet the thresholds and are in Table VI.

Argument *Q* did not occur in the subset (and was only used in 1% of student answers). The authors therefore decided that this code was not applicable to the data from the calculus-based cohort.

This ultimately refined code book was used to code all 944 responses of the calculus-based cohort. It is also this version of the code book that is printed in the paper.

To give insight into student answers' giving rise to disagreement among raters, we discuss an example. Student D answered the item about an inversely proportional function in the mathematics context starting from a symbolic representation with the following explanation:

> *"The relation between* ***x and y*** *is* ***inversely proportional****, so* ***if x increases, the value of y will decrease*** *(a is a positive constant). Because x is not just negative, we divide a by it, the* ***graph will have a parabolic shape.*** *Moreover,* ***there is no starting point b given****, so the answer is g.*

This was coded: $G1, G2, D$, and *S2* by the first rater. However, the second rater did not rate it as *G1*. In the discussion, the team agreed that the excerpt *there is no starting point b given* is categorized as a *G1* argument, because it refers to the intersection with an axis, and based on this argument the student chooses a graph without an intersection with the axes.

TABLE VII. Prevalence of arguments, expressed as percentages. For the algebra-based cohort $N = 1224$ and for the calculus-based cohort $N = 944$.

Argument	*G1*	*G2*	*G3*	*S1*	*S2*	*D*	*I*	*P*	*Q*
Algebra	48	28	14	34	24	45	15	23	6
Calculus	68	48	40	32	49	34	3	15	1

C. Comparing the prevalence of arguments between the two cohorts

Table VII shows the prevalence of arguments (expressed as the ratio of how many times an argument is used in an answer divided by the total number of answers) for both cohorts. The percentages in the table represent only the prevalence, not whether the argument was used in a correct or incorrect way.

There are big quantitative differences between the algebra-based and calculus-based cohorts: the students in each cohort clearly use different arguments. In the next sections we investigate the effects of context, representation, and function type on students' use of arguments.

D. Main effect of context on choice of arguments

To study the effect of context on argument choice, we again use a GEE analysis. For every type of argument we performed the analysis with "occurrence" as the dependent variable and context, direction of translation, and function type as the independent variables. Table VIII shows the prevalences of the arguments for each context in percentages. Percentages with the same superscript are not significantly different.

The analysis shows that there is a main effect of context on choice of arguments for almost all arguments, in both cohorts. Despite the differences in the percentages, we observe remarkably similar patterns.

Pairwise analysis shows that graphical arguments ($G1, G2$, and $G3$) and arguments based on slope (*S1*) are chosen more in mathematical items than in items in a physics or kinematics context. Arguments in which the relation between variables is explicitly named or written in symbolic form (*S2*) occur most in physics items. In the physics and kinematics items, argument *D* involving the covariance of variables is used more often.

Another interesting result is the use of argument *P* (use of physical knowledge). As expected, argument *P* is used least in a mathematics context, but the difference in prevalence between physics and kinematics contexts is also remarkably big. In both cohorts, students used physics knowledge in their answers more often in kinematics than in physics contexts (two to even three times as often).

TABLE VIII. Prevalence of arguments per context, expressed as percentages. N is the number of items per type of context. Significant differences in the use of arguments between the different contexts for both educational backgrounds are represented by superscripts. Prevalences sharing the same superscript are not significantly different from each other.

	Algebra			Calculus		
	M	P	K	M	P	K
	($N = 424$)	($N = 416$)	($N = 384$)	($N = 308$)	($N = 324$)	($N = 312$)
$G1$	57^a	46^b	41^b	82^a	60^b	62^b
$G2$	38^a	26^b	18^b	70^a	40^b	36^b
$G3$	23^a	12^b	6^c	62^a	32^b	27^b
$S1$	48^a	26^b	27^b	52^a	18^b	30^b
$S2$	21^b	34^a	15^b	38^b	61^a	$48^{a,b}$
D	28^b	55^a	52^a	16^b	44^a	42^a
I	22^a	12^b	10^b	4^a	1^a	4^a
P	5^c	20^b	44^a	0^c	10^b	36^a
Q	7^a	4^a	6^a		not applicable	

E. Main effect of the direction of translation on choice of arguments

Table IX shows the difference in the use of arguments between the two directions of translation. The significant differences (in bold in Table IX) are based on a pairwise comparison in the GEE analysis.

Qualitatively the cohorts have more commonalities than differences. It is remarkable that in ten out of twelve cases, there is no difference in prevalence of graphical or symbolic arguments between translating $G \to S$ or $S \to G$. By contrast, the students refer to the covariation of variables (D), and use a numerical approach (I) significantly more often when translating $S \to G$.

It is remarkable that physics-based arguments are used significantly more often by the calculus-based cohort when translating $G \to S$; the difference is nearly significant, but in the opposite direction, for the algebra-based cohort. We cannot readily explain this observation.

TABLE IX. Prevalence of arguments per direction of translation, expressed as percentages. N is the number of items per type of direction. Significant differences in the use of arguments between the different directions are printed in bold.

	Algebra		Calculus	
	$G \to S$	$S \to G$	$G \to S$	$S \to G$
	($N = 597$)	($N = 617$)	($N = 476$)	($N = 468$)
$G1$	**42**	**54**	67	68
$G2$	29	26	52	45
$G3$	15	13	37	44
$S1$	34	34	33	32
$S2$	23	24	51	47
D	**36**	**53**	**30**	**39**
I	**10**	**19**	1	5
P	19	26	**23**	**8**
Q	6	5	not applicable	

F. Main effect of function type on choice of arguments

For this variable, the results become much more complicated. Because of readability we do not share the prevalences and effects for the different function types.

The only clear trend that is observable in our analysis is that the types of arguments students use when they solve items on the inversely proportional function seem to differ from the other function types. This means that all arguments are always used most often or least often in items on the inversely proportional function.

We do not go more into detail in this analysis because of the complexity, but we hope to explore these results further in a next study.

V. DISCUSSION AND IMPLICATIONS FOR TEACHING

In this study we investigated the influence of context, direction of translation and function type on students' ability to link graphical and symbolic representations. One group of 306 students from an algebra-based course and a group of 236 students from a calculus-based course participated in this research.

A. Accuracy

The quantitative data discussed in Sec. III allows us to answer how accurate students are in translating between graphical and symbolic representations of a number of functions frequently encountered in physics. Globally, just over half of the responses given by students in our algebra-based cohort were correct, compared to 80% of the responses by students in the calculus-based cohort. In both cases, correct responses were about 15% more common in a mathematical context. These results suggest strongly that both in calculus-based and in algebra-based physics courses, teachers should not assume that all of their students possess representational fluency: this appears unsafe for about one-fifth and one-half of the students, respectively.

Regarding context, it is in line with our expectations and many results from the literature [8] that items in a mathematics context are solved better than items in physics and kinematics contexts. The finding that there is no difference between physics and kinematics contexts is somewhat surprising. Kinematics is typically taught with a strong graphing component; the physics topics in the test items rarely, if ever. It is reasonable to infer that the way graphs are used in kinematics does not actually support students' understanding of the corresponding equations. (The data of course also admit the possibility of perfect transfer from kinematics to physics contexts, but perfect transfer has never been observed.)

There are 8% more correct responses on items asking students to translate a graphical representation to a symbolic representation than the other way around. Regarding function type, we observe that items on inversely proportional functions are solved worst by the students in all contexts, while the results of the other three functions do not significantly differ from each other. The inversely proportional function was the only nonlinear function in the test. This type of function may be harder for students to recognize. From the study of De Bock *et al.* [29] we can link this result to the tendency to inappropriately connect proportional models to nonproportional situations. De Bock *et al.* also observed this tendency for items on affine functions, but we do not observe a difference between proportional and affine functions.

There are no interaction effects between context and direction of translation, their effects are additive.

For the interaction between context and function type, the effect of context (mathematics items seem to be easier for students) is mainly observable for the more "difficult" function types (decreasing affine and inversely proportional). We do not observe a context effect for the "easier" function types. Students are better at solving items on these function types, independent of the context.

For the interaction between function type and translation, we see that for increasing affine and inversely proportional functions, it is easier for students to start from a graph and identify the correct equation than to start from the equation to identify the graph. For the other function types, we see no such effect.

B. Arguments

The second part of the study focuses on the question "How are the arguments students use when explaining their response affected by context, translation or function type?" Using a code book that is developed bottom-up from the data, we group the arguments in three main categories: graphical, symbolical, and other types of arguments. The same categorization scheme is used in both educational contexts, and we again observe similar trends in both data sets. The analysis reveals an effect of context for almost all types of arguments. We will discuss the two most interesting ones.

The use of physical knowledge (P) is much more prevalent in the kinematics context than the physics context, by a factor of two to three, while the prevalence of graphical arguments is essentially the same. This is surprising, since both cohorts of students would have spent much more class time studying kinematics, including both representations, than any of the physics contexts, which likely only included the symbolic representation. Perhaps paradoxically, greater prior exposure to both representations in kinematics does not correlate with greater accuracy or greater use of graphical arguments, but to more frequent explicit use of physics arguments.

The second trend is the overall difference in argument use between mathematics on the one hand, and physics and kinematics items on the other. From Table VIII it is clear that, when there is an effect of context on argument use, the use of an argument in a mathematics context differs almost always significantly, indicating that students experience this context as different in nature and requiring different solution methods. This is in line with the idea that "math in math" and "math in physics" are like different languages, that physicists make meaning with mathematics in a different way than mathematicians [5,12].

Moreover, in our test mathematics items can be seen as context-free while physics and kinematics have a context loaded onto that same question. If we look at all the available arguments students can use to solve the items, we can conclude that all arguments can be used in items with context, but not all arguments are suited for context-free items (e.g., P). Although students have more arguments or strategies available to solve the items with context, we see that they score worse on these items.

When analyzing the effect of translation on argument use, we expected that graphical arguments would be used more in items asking for a translation starting from a graph and that symbolic arguments would be used more in items asking for a translation starting from an equation, but this does not show in our results. Only argument *G1* in the results from the algebra-based cohort shows a main effect of translation. For the other types of graphical and symbolic arguments there is no directional dependency in translating. We identify two possible reasons for this. A first reason is the fact that we ask students to translate between two representations, so in both types of items there are graphical and symbolic representations present in the problem statement. A second issue is that the line between graphical and symbolic arguments is sometimes vague. You could for example argue that category D, covariance of variables, could have been a symbolic argument as well. We have kept these limitations in mind while interpreting the results.

Finally, we investigated the effect of function type on the use of arguments. After looking deeper into function type effects by doing a pairwise comparison it is clear that these effects are a lot harder to interpret. The only clear trend observable in our analysis is the difference in the arguments used for items on inversely proportional functions. This indicates a different approach to solving the item in the case of the only nonlinear function in the test. Students solve the items on inversely proportional functions worst, and explain their answer by using different arguments. It might be important to keep in mind when teaching that students approach problems on inversely proportional functions differently, solve them in a different way and fail more often to find a correct answer.

C. Qualitative comparison of the algebra-based and calculus-based cohorts

Even though our data were collected in two different settings, with students from different backgrounds and different national contexts, we observe that qualitatively the trends in the results are very similar. The calculus-based cohort performed better than the algebra-based cohort, which one could expect considering their difference in experience and background and from literature [7,18,25]. The fact that they did so, demonstrates (concurrent) validity of our instrument.

For context, in both cohorts, we observed that mathematics items were solved significantly better than physics and kinematics items. Students in the calculus-based course answered 90% of the items set in a mathematics context correctly, which suggests mastery in this area, and about 75% of the items in the physics and kinematics contexts, which suggests that mathematical mastery is not enough to solve physics questions. In the algebra-based cohort just under two-thirds of the items in a mathematics context were correct. This cohort therefore did not display mathematical mastery, and they answered 50% of the items in physics and kinematics contexts correctly. Together these results suggest that improved mathematical skills would benefit students, but that other approaches are needed to help students achieve the same success rate in physics and kinematics contexts. For the direction of translation, both cohorts were better at solving items starting from a graphical representation. Lastly, for function type, both cohorts has significantly lower accuracy for items on inversely proportional functions.

In the second part of the study we focused on argument use when students explain their choice of answer. Again, there are similar trends observed. However, it becomes more difficult to compare the effects. In this study, we cannot perform a quantitative comparison between the two cohorts because the differences between the cases were not quantified, nor controlled.

D. Future research and limitations of the study

As a next step we would like to further investigate student argument use and especially how this is linked to the accuracy of their answer. It would be interesting to identify if the explanation of the student as a whole, so the sum of all the used arguments, is correct or not and to link that to the accuracy of the chosen multiple-choice alternative.

There were several limitations to the study design and implementation that should be discussed. First, the choice of the institutions and participating students was not deliberate, but rather of a practical nature. We cannot compare the results of both cases quantitatively because of the many differences. However, this is not the aim of the study and should therefore not be a problem. Second, it is important to stress once more that the code book used in both cohorts was not exactly the same. The version used for the calculus-based cohort had the same argument categories, but was refined and extended with some specific examples. Third, the discussion of the argument use had some limitations. We identified many different arguments, some with a higher prevalence than others. Low prevalences for arguments can result in unreliable results from Cohen's kappa statistics, as we stressed in the discussion of these results. In this study we limited our analysis to the prevalence of the separate arguments, and no information on the correct use of the arguments is given. It would be interesting to also look at the combination of arguments, but further research is necessary to look into this.

ACKNOWLEDGMENTS

We thank Professor Johan Deprez (KU Leuven), Professor Martin Goedhart (University of Groningen), and Dr. Laurens Bollen (KU Leuven) for reading our manuscript and sharing their constructive thoughts with us. Also, we would like to thank all the participating students in Leuven and Dublin. This publication was made possible through funding support of the KU Leuven Fund for Fair Open Access.

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

[2] S. E. Ainsworth, P. A. Bibby, and D. J. Wood, Analyzing the costs and benefits of multiple-representational learning environments, in *Learning with Multiple Representations*, edited by M. W. v. Someren (Pergamon, New York, 1998).

[3] R. Even, Factors involved in linking representations of functions, J. Math. Behav. **17**, 105 (1998).

[4] A. A. Nistal, W. Van Dooren, G. Clarebout, J. Elen, and L. Verschaffel, Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: a critical review, ZDM Int. J. Math. Educ. **41**, 627 (2009).

[5] *Science & Education*, edited by R. Karam (Springer, Switzerland, 2015), Vol. 24.

[6] 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).

[7] 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).

[8] 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).

[9] 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** (2018).

[10] 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).

[11] 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).

[12] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, Sci. Educ. **24**, 561 (2015).

[13] D. L. McKenzie and M. J. Padilla, The construction and validation of the test of graphing in science (TOGS), J. Res. Sci. Teach. **23**, 571 (1986).

[14] H. M. Brasell and M. B. Rowe, Graphing skills among high school physics students, School Sci. Math. **93**, 63 (1993).

[15] A. Bell and C. Janvier, The interpretation of graphs representing situations, For the Learning of Mathematics **2**, 34 (1981).

[16] J. Clement *et al.*, *Adolescents' Graphing Skills: A Descriptive Analysis* (ERIC, 1985).

[17] 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).

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

[19] 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).

[20] F. M. Goldberg and J. H. Anderson, Student difficulties with graphical representations of negative values of velocity, Phys. Teach. **27**, 254 (1989).

[21] R. K. Thornton and D. R. Sokoloff, Learning motion concepts using realtime microcomputer-based laboratory tools, Am. J. Phys. **58**, 858 (1990).

[22] R. J. Beichner, The impact of video motion analysis on kinematics graph interpretation skills, Am. J. Phys. **64**, 1272 (1996).

[23] E. MH Billings and D. Klanderman, Graphical representations of speed: obstacles preservice K-8 teachers experience, School Sci. Math. **100**, 440 (2000).

[24] I. S. Araujo, E. A. Veit, and M. A. Moreira, Physics students performance using computational modelling activities to improve kinematics graphs interpretation, Comput. Educ. **50**, 1128 (2008).

[25] L. Bollen, P. van Kampen, C. Baily, and M. De Cock, Qualitative investigation into students use of divergence and curl in electromagnetism, Phys. Rev. Phys. Educ. Res. **12**, 020134 (2016).

[26] A. Acevedo 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).

[27] W.-M. Roth and K. Tobin, Cascades of inscriptions and the representation of nature: how numbers, tables, graphs, and money come to represent a rolling ball, Int. J. Sci. Educ. **19**, 1075 (1997).

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

[29] 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).

[30] 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).

[31] P. B. Kohl and N. D. Finkelstein, Effects of representation on students solving physics problems: A fine-grained characterization, Phys. Rev. ST Phys. Educ. Res. **2**, 010106 (2006).

[32] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.15.020113 for complete collection of developed test items.

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

[34] V. Otero and D. Harlow, *Getting Started in Qualitative Physics Education Research* (2019).

[35] A. J. Viera and J. M. Garrett, Understanding interobserver agreement: the kappa statistic, Fam. Med. **37**, 360 (2005).

[36] J. Clement, The Concept of Variation and Misconceptions in Cartesian Graphing, Focus Learn. Probl. Math. **11**, 77 (1989).

[37] S. Ceuppens, L. Bollen, J. Deprez, W. Dehaene, and M. De Cock, 9th grade students' understanding and strategies when solving $x(t)$ problems in 1D kinematics and $y(x)$ problems in mathematics, Phys. Rev. Phys. Educ. Res. **15**, 010101 (2019).