# Design and validation of a test for representational fluency of 9th grade students in physics and mathematics: The case of linear functions 

Stijn Ceuppens* ${ }^{*}$<br>KU Leuven, Department of Physics and Astronomy, Celestijnenlaan 200c box 2406, 3001 Leuven, Belgium<br>Johan Deprez<br>KU Leuven, Department of Mathematics, Celestijnenlaan 200b box 2400, 3001 Leuven, Belgium<br>Wim Dehaene<br>KU Leuven, ESAT-MICAS, Microelectronics and Sensors, Kasteelpark Arenberg 10 box 2443, 3001 Leuven, Belgium<br>Mieke De Cock ${ }^{\dagger}$<br>KU Leuven, Department of Physics and Astronomy, Celestijnenlaan 200c—box 2406, 3001 Leuven, Belgium

(Received 1 February 2018; published 10 August 2018)


#### Abstract

This study reports on the development, validation, and administration of a 48-item multiple-choice test to assess students' representational fluency of linear functions in a physics context (1D kinematics) and a mathematics context. The test includes three external representations: graphs, tables, and formulas, which result in six possible representational transitions between them. Moreover, four linear function types are included: negative $y$ intercept and positive slope, zero $y$ intercept and positive slope, positive $y$ intercept and negative slope, and positive $y$ intercept and positive slope. The test is administered to 385 students aged $14-15$ in the 9 th grade from 13 schools in Flanders (Belgium) and—after validation-is analyzed by means of generalized estimating equations (GEE). The results show a significant main effect for all design factors and a significant interaction effect between representational transition and function type, as well as additional interaction effects with the gender of the respondents. Furthermore, representational transitions which include a formula prove to be significantly more difficult; in particular for the directly proportional function type, the transition to a formula stands out from the analysis. Mean accuracies in physics are significantly lower compared to mathematics. Function types with negative values for either $y$ intercept or slope also result in significantly lower mean accuracies indicating the difficulty students have with negative numbers in linear functions. A distractor analysis of the incorrect answers chosen by the students reveals three distinct dominant errors: concept switching, sign switching, and switching of the directly proportional function type and the function type with positive $y$ intercept and positive slope.


DOI: 10.1103/PhysRevPhysEducRes.14.020105

## I. INTRODUCTION AND LITERATURE

## A. Physics and mathematics

We report on the design and validation of a test to study representational fluency in physics and mathematics on the specific topic of the uniform linear motion in physics and

[^0]Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.
linear function problems in mathematics. The motivation for the test covering these two fields is twofold.

First, mathematics and physics have always been deeply linked, although not always in the same way. As described by de Ataide and Greca [1] the role of mathematics in physics historically evolved in three stages: from a way to describe analogies between physical objects in the real world and structures in an ideal world, to a mere language or a convenient tool to describe reality, and, finally, to a primary driver of new physical knowledge. This evolution came quite naturally with the ever-growing complexity of physical knowledge and the need for a consistent framework and unified theories to describe, to explain, to predict, and to understand physical reality. Physics has become inseparable from mathematical modeling and mathematics
is no longer simply being "applied" to physics as stressed by Kjeldsen and Lützen [2]. This view is further supported by Redish and Kuo [3] who noted, "we not only use math in doing physics, we use physics in doing math." So, the relation between these two fields is no longer a one-way street but has become a two-way interaction where mathematics can be a driver for new physics and vice versa. Mathematics has essentially become a language loaded with physical meaning plus reasoning to physics. According to Redish and Kuo [3] physicists appear to use a different dialect of the "mathematical language." This means that students must not simply learn mathematics, but also need to learn the specific dialect used in physics. They must learn to link the pure mathematical language to a physical context.

Second, good mathematical knowledge and skills are necessary for good problem-solving ability in physics, but are not sufficient on their own [3-5]. This again indicates that there is more to it than just knowing mathematics and knowing physics. Linking these two fields is also part of the requirement for good problem-solving ability. This, however, is a persistent problem for students in all educational levels as is shown by several studies. Pollock, Thompson, and Mountcastle [6] compared questions related to the first law of thermodynamics and equivalent context-less questions in mathematics. They provided evidence that students' understanding was isolated in each field. Also Christensen and Thompson [7] studied slope and derivative in what they call "physics-less physics questions." These are questions that look like physics questions but are stripped of their physical context. They found that students have difficulties conceptualizing mathematical questions portrayed as physics questions, indicating a lack of transfer of knowledge from mathematics to physics. Planinic et al. [8] compared students' understanding of slope and area under a graph in mathematics, physics, and other contexts. They found that mathematical questions were easier and that physics remains a more difficult context. Wemyss and van Kampen [9] studied students’ interpretations of linear distance-time graphs and found that the ability to determine the slope of a $y-x$ graph and having a qualitative understanding of distance-time graphs was insufficient to determine the speed on a distance-time graph.

The essential insight from the bulk of these studies is that good understanding in one field does not necessarily imply good understanding in the other field. Furthermore, the literature suggests that students' abilities in physics are not independent of their abilities in mathematics, nor are they independent of their abilities to link these two fields and to blend knowledge from both. It is clear then that the blending of these fields must be an area of focus in education and, consequently, requires various means for assessment and gathering data to improve insight therein. For this purpose we use a "three-pronged approach" where students' abilities are separately tested in a first
context, in a-as isomorphic as possible-second context and then compared between contexts, which is in line with the approach in several other studies [6,8-10].

## B. The influence of the representational format

Besides the context in which a problem is formulated, another key aspect that influences the success rate and the solution strategy is the representational format in which the question is asked [11]. The representations used most often in scientific disciplines are formulas, graphs, tables, diagrams, and verbal descriptions. Each of these can have many variants and each difference can influence the final solution and the strategy to obtain it. Various studies demonstrate these effects. Meltzer [11] studied students' responses on two of Newton's third-law questions, one in a verbal representation and the other in a diagrammatic representation, and found consistent differences in success rate and in the errors made. Additionally, he used questions in verbal, diagrammatic, mathematical or symbolic, and graphical representations and also found evidence of an interaction effect between gender and representation. Similarly, Kohl and Finkelstein [12] concluded that the success rate in solving physics problems shows significant differences for near-isomorphic questions presented in a mathematical, pictorial, graphical, and verbal representation. A second observation from their study is that allowing students to choose the representation in which they solve a given physics problem, can for some students increase and for others decrease the success rate. Acevedo Nistal, Van Dooren, and Verschaffel [13] studied students' flexibility when choosing a representational format (formula or table) to solve linear function problems in mathematics in secondary education. They found that students have an overwhelming preference to use formulas, which increases with grade. Even when tables are expected to be more effective, students most often choose formulas and achieve greater accuracies when compared to using tables. A study from De Cock [14] also confirmed such effects by investigating the strategy students chose when presented with a question in either a verbal, pictorial, or graphical representation in a physics context. By analyzing students' explanations in support of their answer, she was able to show that details in the used representational format influenced the chosen problem solving strategy. Lastly, De Bock, Van Dooren, and Verschaffel [15] conducted a study on the effect of representations in a more general context. Students had to model a textual description with either a graphical, tabular, or formulaic representation of the situation. Four different mathematical models were studied: proportional, inversely proportional, affine with negative slope, and affine with positive slope. Their analysis identified a significant interaction effect between the mathematical model and representation, meaning that the representational format can have a facilitating effect for a particular model but an inhibiting effect for another.

For both the proportional model and the affine model with negative slope, formulas resulted in significantly lower scores compared to tables and graphs. For the inversely proportional model though, results from formulas were significantly better compared to those from tables and graphs. For the affine model with positive slope, graphs resulted in significantly higher scores compared to tables and graphs. These studies demonstrate the important effect that using different representations can have. This is a strong argument to support students in learning different representations and the links and transformations between them.

## C. Representational fluency

There are several representational abilities defined and categorized in literature. A recent insightful summary is given in Deliyianni et al. [16] as well as in Hill, Sharma, O'Byrne, and Airey [17]. In the work presented here we focus on representational fluency, defined as the ability to interpret and construct representations, and the ability to translate between them. This is in accordance with earlier work by one of us [14], from Nistal et al. [18], from the more general work of Ainsworth, Bibby, and Wood [19], and from Even [20] on linking representations. Our definition of representational fluency is very specific and our focus here is on the ability to translate between representations. We are interested in how fluent students are in translating (identifying) the same concept-in this case a type of linear function-to (in) a different representation. This definition is closely related to interrepresentation flexibility, which is the handling of between-representation transformations and combines both recognition and conversion competences, as is defined by Deliyianni et al. [16]. The processes students use to perform these transitions are not part of this study, but we do reflect on this in Sec. V.

Studies such as those of Ainsworth, Bibby, and Wood [21], Duval [22], and Kirsh [23] specifically identify the transition between representations as a key task in learning and problem solving; Duval even calls it the true challenge of mathematics education. Furthermore, Ainsworth, Bibby, and Wood focus heavily on transitions between multiple external representations (MERs) and proposes a functional taxonomy of MERs [24] which they later incorporates into the DeFT (design, functions, tasks) framework for learning with MERs [25]. They identify three key functions: (i) complementary roles; (ii) constrain interpretation; (iii) construct deeper understanding. In this framework they make clear that a multirepresentational environmentwhich should include representational fluency-can be of great benefit to the learning process.

## D. Literature on representational fluency, linear functions, and kinematics

In his study on fractions in mathematics, Niemi [26] showed that the level of representational knowledgewhich he measured by the ability to identify correct
alternative representations of a fraction given in a particular representation-is predictive for performance in problem solving, justification, and explanation tasks. When comparing novices and experts in physics several researchers have shown that there are important differences in representational abilities between these groups. A study by Chi, Feltovich, and Glaser [27] indicated that representational features are more distracting for novices than for experts, resulting in the inability to identify the underlying concept to solve the problem. They also found that when asked to categorize a selection of physics problems, novices sorted them by diagrammatic format and diagrammatic similarity whereas experts did so by the underlying physical concepts. Additionally, Kohl and Finkelstein [28] learned that-when asked to solve a physics problem-experts are more flexible in the initially chosen representation to work with and are likely to switch quicker between representations than novices.

In their work, De Bock, Van Dooren, and Verschaffel [15] also reported on a second study. In this study they investigated linking graphical, tabular, and symbolical (formulas) representations for four different mathematical models: proportional, inversely proportional, negative affine functions, and positive affine functions. Each question described a general context, included one of these models as either a graph, a table, or a formula and asked to identify the matching option from the possible multiplechoice options. These choice options were presented as a graph, a formula, or a table, but were always different from the representation used in the questions. They showed that the number and nature of errors strongly depend on the representational transition. Students performed best when switching or translating between graphs and tables (in both directions), and worst when switching or translating between formulas and graphs. They concluded that representations that contain concrete function values are easiest to transition from and to. Additionally they found students to have the most difficulty with models with decreasing functions. Their most important finding was the presence of a significant interaction effect between the representational transition and the type of function (model). They found that starting from a formula was the most difficult and that going to a table was the easiest for all function types, but that for affine functions with negative slope the accuracies were far lower when going to a graph or to a formula. In addition, for inversely proportional functions they found that when switching to or from a formula the transition is also far more difficult. They obtained the highest accuracies for proportional functions and affine functions with positive slope. In general they found that functions with a negative slope are the most difficult to tackle, similar to the results from other studies [8,9]. Research on student understanding of the $y$ intercept of linear functions by Davis [29] showed that using terminology such as "starting point" for the $y$ intercept in real-world contexts (e.g., a fall from a certain
height) negatively impacted their performance on translating between symbolic, graphical, and tabular representations. This indicates the difficulties students have with transferring knowledge about the $y$ intercept between representations and contexts similar to results from Schoenfeld et al. [30].

In the specific context of kinematics in physics such problems have been studied for a long time, with specific attention to position-time graphs in kinematics $[7,8,10$, 31-36]. Of note is the TUG-K test from Beichner [35] which was specifically designed to assess student knowledge of graphs in kinematics, as well as the work by Planinic et al. $[8,10]$. They also studied linear graphs in mathematics, kinematics, and contexts other than physics and found that students in secondary education had more difficulty with the physics questions and specifically with negative velocity. University students, though, had closed the gap and achieved similar results on questions about slope, independent of context. This suggests they were able to successfully transfer their mathematical knowledge.

Representational fluency is often indirectly included in other tests that use various different representations, but it is rarely the goal of the test itself. One study we explicitly mention, is that of Hill et al. [17] that reports on the development and validation of the representational fluency survey (RFS) for university physics students, which is a broad scale multiple-choice survey composed of previously validated items for other tests and newly written items in words, graphs, equations, and diagrams. In contrast with the very broad scale RFS test, our work focuses on linear relations and the differences between physics and mathematics specifically.

## E. Research questions

The main goal of our study is to construct and validate a multiple choice test on representational fluency in linear function problems in mathematics and physics. Thereby, we adhere to the definition of representational fluency given above. However, one could equally well use the term representational recognition for our study, which is the ability to recognize the same concept in multiple representations. We aim to set up a similarly structured test as that of De Bock, Van Dooren, and Verschaffel [15] but with a different selection of function types (models), a nearisomorphic comparison between physics and mathematics, and a different age group. Their respondents had typically finished three years of nonuniversity higher education; our respondents are aged $14-15$, still in secondary school, and have only recently studied linear functions in mathematics and 1D kinematics in physics. We focus on these first order relations only. Variations are made through different sign combinations of the $y$ intercept and slope in mathematics and on initial position and velocity of uniform linear motion in physics. These types of linear relations are key topics in science education for our target students.

With this test structure we aim to answer the following research questions:
(1) How does student representational fluency between graphs, tables, and algebraic formulas of linear functions differ between mathematics and physics (1D kinematics)?
(2) How does student representational fluency between graphs, tables, and algebraic formulas differ for different combinations of slope sign and $y$-intercept sign in linear function problems in mathematics and physics (1D kinematics)?
(3) What are the significant interaction effects between representational transition, context, and linear function type (combination of slope and $y$-intercept sign)?
In addition to answering these research questions the test can also be used as a tool to compare and contrast different teaching approaches for representations and determine which approach is more successful.

To answer these questions, we first describe the test design in Sec. II, followed by a description of the test procedure in Sec. III and results in Sec. IV. Next, we discuss the validation, the results, and future research in Sec. V, in which we also highlight some guidelines for teaching. We summarize the main findings in Sec. VI.

## II. TEST DESIGN

For our purposes the test needs to incorporate the three aspects of interest: different representational transitions, different linear function types, and different contexts. The full test is translated to English and available for download as Supplemental Material [37].

## A. Representational transitions

Test items are formulated in one of the three representations: tables (T), graphs (G), and algebraic formulas (F). This choice is motivated by their abundant use by experts in both mathematics and physics, as well as by the many problems students experience when using graphs and the need they often feel [38] to use formulas to solve problems. The answer possibilities in the multiple-choice items are provided in either of the two remaining representations. This means students are forced to make the transition requested in the item, thus allowing the test to measure representational fluency and not representational flexibility, which would assess the representational choices students make. This results in six possible representational transitions: table to graph (TG), graph to table (GT), formula to table (FT), table to formula (TF), formula to graph (FG), and graph to formula (GF).

## B. Linear function types

Linear functions can be categorized in six different types when distinguishing between all combinations of negative


FIG. 1. Graphical representation of the six possible answer options. Full lines are the function types under study, dashed lines are buffer distractors. Abbreviations denote the sign ( $N$ for negative, $Z$ for zero, and $P$ for positive) of the $y$ intercept and the slope, respectively.
or positive slope and negative, zero, or positive $y$ intercept as shown in a graphical representation in Fig. 1. To limit the number of questions of the test and maximize the relevance of these possible combinations for kinematics, four out of six are selected to be studied here: negative $y$ intercept with positive slope (NP), zero $y$ intercept with positive slope (ZP), positive $y$ intercept with positive slope ( PP ), and positive $y$ intercept with negative slope (PN). For each multiple-choice question students are allowed to choose their answer from the full set of six possible combinations of $y$ intercept and slope, thus also including the two remaining possible combinations ( NN and ZN function types) as buffer distractors, which are marked with dashed lines in Fig. 1.

## C. Contexts

The test consists of two parts that are designed to be as isomorphic as possible. The first part is in a mathematics context, meaning questions provide information about the relation of $y$ as a function of $x$ and offer no additional context or interpretation of the dependent or independent variable. The second part is in a physics context (1D kinematics). The questions describe the uniform linear motion of a cyclist and provide information about the position $x$ as a function of the time $t$. Depending on the representation, implicit or explicit information is provided about the sign of the initial position $x_{0}$ and the velocity $v$, where $v$ is interpreted as the slope of the curve and can thus have a negative sign. The negative sign of velocity, particularly in graphs, has been shown to be problematic for many students [34] and is an extra point of interest in the test. Although the physics context and content in the physics questions might seem very low to some, we expect that the differences with the mathematics questions are meaningful and significant enough for students and that they will trigger a different mindset in which to solve the questions and consequently make different errors. When comparing this to what Christensen and Thompson [7] labeled "physics-less physics questions," it becomes clear that our questions do not have the physics context stripped
off, thus warranting the designation of true "physics items." Figure 2 provides examples of four items from the test, including questions P9 and M9-the physics (P) and the mathematics (M) version of question 9-to illustrate the contextual differences.

## D. Other design choices

Some additional design choices are made to keep the test consistently structured and to eliminate any superfluous distracting information that might result in an incorrect answer unrelated to the essential concept under investigation. These are as follows:
(i) All chosen linear function types must be clearly visually distinguishable from each other in the graphical representation.
(ii) Mainly the first and fourth quadrants in the graphical representation are shown to avoid the use of negative times in a kinematics context. A small section of the second and third quadrants are also shown. In mathematics, these small sections are used to make clear that the domain of the function value is $]-\infty,+\infty[$. In the kinematics context negative times are discarded and the domain for the time variable $t$ is $[0,+\infty[$.
(iii) The root is not present for any of the graphs except for the directly proportional one to prevent possible confusion between the $y$ intercept and the root.
(iv) The $y$ intercept is explicitly present in both graphs and tables.
(v) In the tables, the values of the independent variable are always the same: $0,1,2,3$, and 4 . The inclusion of value 0 facilitates the deduction of the $y$ intercept and the use of equidistant steps facilitates deduction of the slope. A random rotation between multiples of $3,4,6$, and 7 (positive and negative) is used for the slope and the velocity since these still result in small numbers, thus avoiding potential problems with large numbers, and are not the easiest number either, such as 1,2 , or 5 .
(vi) All formulas are represented in the same general format accompanied by declarations about the signs of the slope and the $y$ intercept. For physics the terms are ranked from lowest degree to highest degree, for mathematics from highest degree to lowest degree.

## E. Test structure

All the previously listed choices result in a 48-item multiple choice test with a block of 24 mathematics questions and a block of 24-as isomorphic as possiblephysics questions. For each representational transition there are a total of 8 questions, and for each function type there are 12 questions. To prevent a possible effect from the order of the contexts-first physics then mathematics or the other way around-about half of the students received the

P1 A cyclist is cycling at constant velocity on a straight road. The table describes the relationship between the position $x$ (in meters) and the time $t$ (in seconds):

M21 The variable $y$ is dependent of the variable $x$ according to a first degree equation. The graph describes this relationship between $y$ and $x$ :


One of the equations describes the same relationship between $y$ and $x$. Which one?
a) $y=a x+b$ with $a>0$ and $b>0$.
b) $y=a x+b$ with $a<0$ and $b=0$.
c) $y=a x+b$ with $a>0$ and $b<0$.
d) $y=a x+b$ with $a>0$ and $b=0$.
e) $y=a x+b$ with $a<0$ and $b<0$.
f) $y=a x+b$ with $a<0$ and $b>0$.
(b)

M9 The variable $y$ is dependent of the variable $x$ according to a first degree equation $y=a x+b$ with $a>0$ and $b<0$.
One of the equations describes the same relationship $y$ and $x$. Which one?

| a) |  |
| :---: | :---: |
| $t$ <br> (in s) | $x$ <br> (in m) |
| 0 | 0 |
| 1 | 7 |
| 2 | 14 |
| 3 | 21 |
| 4 | 28 |






(c)

| b) |  |
| :---: | :---: |
| $x$ | $y$ |
| 0 | 0 |
| 1 | 7 |
| 2 | 14 |
| 3 | 21 |
| 4 | 28 |$\quad$| c) |  |
| :---: | :---: |
| 0 | $y$ |
| 0 | 42 |
| 1 | 35 |
| 2 | 28 |
| 3 | 21 |
| 4 | 14 |



| f) |  |
| :---: | :---: |
| $x$ | $y$ |
| 0 | 0 |
| 1 | -7 |
| 2 | -14 |
| 3 | -21 |
| 4 | -28 |

(d)

FIG. 2. Example questions from the test. Question P1 is a physics question in which the TG transition from table to graph has to be made. Question M21 is in the context of mathematics and requires a GF transition from graph to formula. Question P9 and M9 are isomorphic questions concerning the formula to table (FT) transition in the context of physics and mathematics, respectively.
mathematics questions first, followed by the physics questions, the other half received the physics questions first and then the mathematics questions. Within each context block the items have a random order, but that order is the same in both context blocks. Each multiple-choice question offers six possible answers: the six possible sign combinations of $y$ intercept and slope. As previously mentioned, these include
the four function types under investigation and the two buffer distractors which are shown in Fig. 1 in dashed lines. While solving the total of 48 questions students might start to realize the two buffer distractors are never the correct answer. To prevent this as much as possible, the order of the six possible answers are randomized differently for each item, though the symmetry between the mathematics and the
physics test is maintained to keep the two blocks as isomorphic as possible.

## F. Expert revision

To validate the structure and set up of the test we consulted with a member of academic staff with expertise in this field. The feedback praised the very systemic structure of the test, allowing an in-depth analysis of all the variables involved. Advise on slight changes in wording were adopted in the final version. Some concern was expressed about the use of numbers in the tabular representations and the absence thereof in the graphical and formulaic representations. Our aim to avoid the use of numbers wherever possible-thus avoiding potential problems with numbers-was conceived positively, but the inevitable use of numbers in tables does create an asymmetry in the test structure which might benefit or disadvantage certain representational transitions. Also, the expert picked up on our choice to use equidistant numbers for the independent variable in the tables, which includes 0 and has a step size of 1 unit. We were urged to clearly motivate this choice. In addition to the main motivations already given in Sec. II D, this choice was also motivated by the time limit since students would likely require more time to deduce the signs of the $y$ intercept and the slope when numbers are nonequidistant-which also automatically results in larger and likely more difficult numbersand by our goal to question all 6 representational transitions for at least 4 function types and 2 contexts.

The last point of the feedback was about issues with negative velocity in 1D kinematics. In English there is the distinction between the magnitude "speed" and the vector quantity "velocity," while in Dutch, there is only "snelheid" to describe both. In our test, we want to question students about snelheid in graphs with a negative slope, in formulas with a negative coefficient in the first order term and in tables with a negative difference between entries for the position. This means we actually ask for the $x$ component of the velocity and not for the speed. Strictly speaking, when using snelheid in 1D kinematics, we mean velocity as a vector $\left(v_{x} \bar{e}_{x}\right)$. We therefore expect the student to not only give the length of the velocity vector, but also include information on the direction by assessing $v_{x}$, which can be negative. A specific example in which this is of importance is item P15, which is a physics questions with a table to formula transition for a PN function type. The table given in P15 contains positive but decreasing values for the position $x$ similar to the one in option c) in M9 shown in Fig. 2. In our analysis, we consider the velocity in every PN question to be negative, which is important when interpreting and comparing our results.

In addition to the revision by the academic expert, we also asked mathematics and physics in-service teachers to provide feedback on the test items to assure content validity. Two in-service teachers in each field were
addressed. Comments were very positive and the overall difficulty was perceived suitable for the target group. Small suggestions for better wording were adopted. Some remarks were made on the difference in difficulty between decreasing and increasing functions. In particular, the teachers pointed out that students are very likely to have more difficulties with the tables in which negative function values are provided, as well as with the graphs of the function type with negative intercept and positive slopethe ones where an increasing function is shown in the fourth quadrant. In summary, the teachers in both fields expected the most difficulties with the use of negative function values and concluded that the test has a high content validity. Note that for function types with negative intercept, all values in the tables are negative since the choice was made to not include the root-except for the ZP function type.

## III. TEST ADMINISTRATION

## A. Participants

The data used for the validation were gathered from administering the test in early 2017 in Flanders (Belgium) with students from the 9th grade aged $14-15$, who had opted to study sciences in secondary education. These participants were from 13 different schools distributed across Flanders. All these schools provide the same type of education, called "aso education." Their students receive education on a broad number of topics with selected specializations such as sciences, languages, economics, humanities, etc., to prepare them for higher education. These schools voluntarily agreed to have the test administered to students in a learning trajectory with a strong science component. Informed consent from the students' parents was obtained for all 385 students who participated in this study ( 213 males and 172 females). All of these students had received instruction on the topic of linear functions in mathematics and on 1D kinematics in physics to achieve specific learning goals set by government standards. All gathered data were collected and treated anonymously to protect student privacy. Ethical approval for this study was granted by the Sociaal-Maatschappelijke Ethische Commissie (SMEC)-Social and Societal Ethics Committee of KU Leuven.

## B. Test procedure

This test was administered digitally using the Qualtrics ${ }^{\circledR}$ online software platform for digital surveys. The digital administration facilitated the randomization of the test structure at the respondent level, meaning that the odds are very small that any two respondents received the exact same test. This is beneficial to reduce possible learning effects due to the order of the questions and of the answers. Furthermore, the digital nature of the administration allowed for forced responses that forces respondents to select an
answer in order to proceed to the next question. This resulted in a full data set without missing values. When administering this test the following procedure was used: respondents were informed that they would be participating in a research project, that no educational credit could be earned through the participation and that the maximum time available to solve all 48 multiple-choice questions was 45 min . Neither students nor their teachers knew about the contents of the test beforehand, thus excluding any possible influence or teaching to the test that might occur.

## IV. RESULTS

## A. Validation

In this section we present the results from the validation [39]. ${ }^{1}$ We provide a per-item overview of relevant validation quantities in Table I. In this table the items are sorted by representational transition and by function type to facilitate easy interpretation. The actual item order in the test during administration was randomized.

## 1. Internal consistency

To assess the internal consistency of the full test, Cronbach's $\alpha$ is calculated, which results in a value of $\alpha=0.946$, indicating a high internal consistency since it surpasses the 0.9 threshold [40]. To check the effect of each question on the internal consistency, $\alpha$ is also calculated for the case in which each item $i$ is omitted from the test. These per-item results show values between 0.944 and 0.946 , meaning that all items are worth retaining since omitting any item would either keep the value for $\alpha$ or lower it.

## 2. Item difficulty index

For dichotomous items such as multiple-choice items, the item difficulty index $P$ is the proportion of the number $N_{c}$ of correct responses to the total number $N$ of responses given to that item:

$$
\begin{equation*}
P=\frac{N_{c}}{N} . \tag{1}
\end{equation*}
$$

A higher value means an item is less difficult and a lower value means an item is more difficult. This measure can thus be used to provide an overview of the difficulty distribution of all the test items. We present the results and the standard deviation for each item in Table I and show that item difficulty indexes for this test range from 0.45 to 0.90 with fairly large standard deviation and an overall

[^1]TABLE I. The relevant validation quantities for each question $Q$ of the test: item difficulty index $P$, the standard deviation $S D$ for $P$ and the item discrimination index $D$. The items are sorted by representation transition (RepT), function type (FType), and context to simplify interpretation and comparison. Question numbers starting with "P" denote a physics context and those with " M " a mathematics context.

| RepT | FType | $Q$ | $P$ | $S D$ | $D$ | $Q$ | $P$ | $S D$ | $D$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| TG | NP | P1 | 0.85 | 0.36 | 0.37 | M1 | 0.86 | 0.35 | 0.43 |
|  | ZP | P2 | 0.90 | 0.30 | 0.30 | M2 | 0.88 | 0.33 | 0.37 |
|  | PN | P3 | 0.85 | 0.35 | 0.35 | M3 | 0.86 | 0.35 | 0.37 |
|  | PP | P4 | 0.86 | 0.35 | 0.37 | M4 | 0.85 | 0.36 | 0.43 |
| GT | NP | P5 | 0.85 | 0.36 | 0.39 | M5 | 0.85 | 0.36 | 0.46 |
|  | ZP | P6 | 0.90 | 0.31 | 0.29 | M6 | 0.88 | 0.32 | 0.39 |
|  | PN | P7 | 0.85 | 0.35 | 0.42 | M7 | 0.88 | 0.32 | 0.37 |
|  | PP | P8 | 0.85 | 0.36 | 0.43 | M8 | 0.82 | 0.38 | 0.44 |
| FT | NP | P9 | 0.45 | 0.50 | 0.80 | M9 | 0.52 | 0.50 | 0.67 |
|  | ZP | P10 | 0.67 | 0.47 | 0.66 | M10 | 0.72 | 0.45 | 0.62 |
|  | PN | P11 | 0.45 | 0.50 | 0.54 | M11 | 0.51 | 0.51 | 0.65 |
|  | PP | P12 | 0.60 | 0.49 | 0.73 | M12 | 0.65 | 0.48 | 0.67 |
| TF | NP | P13 | 0.45 | 0.50 | 0.60 | M13 | 0.59 | 0.49 | 0.60 |
|  | ZP | P14 | 0.62 | 0.49 | 0.71 | M14 | 0.59 | 0.49 | 0.74 |
|  | PN | P15 | 0.46 | 0.50 | 0.69 | M15 | 0.51 | 0.50 | 0.75 |
|  | PP | P16 | 0.60 | 0.49 | 0.80 | M16 | 0.65 | 0.48 | 0.67 |
| FG | NP | P17 | 0.56 | 0.50 | 0.60 | M17 | 0.58 | 0.49 | 0.67 |
|  | ZP | P18 | 0.67 | 0.47 | 0.62 | M18 | 0.75 | 0.43 | 0.55 |
|  | PN | P19 | 0.51 | 0.50 | 0.67 | M19 | 0.61 | 0.49 | 0.69 |
|  | PP | P20 | 0.69 | 0.46 | 0.70 | M20 | 0.73 | 0.44 | 0.66 |
| GF | NP | P21 | 0.54 | 0.50 | 0.59 | M21 | 0.59 | 0.49 | 0.73 |
|  | ZP | P22 | 0.63 | 0.48 | 0.74 | M22 | 0.64 | 0.48 | 0.72 |
|  | PN | P23 | 0.53 | 0.50 | 0.63 | M23 | 0.56 | 0.50 | 0.72 |
|  | PP | P24 | 0.68 | 0.47 | 0.67 | M24 | 0.73 | 0.45 | 0.65 |
|  |  |  |  |  |  |  |  |  |  |

mean and standard deviation of $0.69 \pm 0.24$, respectively. The first eight questions in the table-the TG and GT transition-have very high item difficulty indexes close to 1 , meaning these transitions are very easy for students. Tests can be categorized into two different types: normreferenced tests (NRTs) and criterion-referenced tests (CRTs) [41]. NRTs usually cover a large domain of learning tasks and aim to compare groups and individuals to groups which is why they need to provide good discrimination and average difficulty-depending on the number of options in multiple-choice tests. Ideally the item difficulty indexes (or average scores) are normally distributed with the bulk roughly between 0.4 and 0.6 . CRTs aim to assess the mastery of a specific topic or an individual's ability in regard to a criterion and serve to diagnose students' difficulties and certify their competencies. The item difficulty index distribution is of lesser importance, but is generally strongly skewed to higher scores for instructed groups (more than 0.8) and strongly skewed to lower scores for uninstructed groups (less than 0.3) [42]. Furthermore, items which are found to be too easy or too
difficult are not omitted from CRTs, whereas they usually are from NRTs. Since our test is designed to assess a specific competency (representational fluency) for a specific topic, it is more CRT than NRT oriented. This means that our results indicate that the item difficulty indexes we have found are arguably well distributed for our purposes.

## 3. Discriminatory power

A test should allow for discrimination between respondents with high ability and respondents with low ability. This is known as the discriminatory power of a test and can be studied at the item level as well as at the full test level. To describe the power of a single test item to discriminate between high- and low-achieving respondents, we calculate $D$, the item discrimination index [40]. To do so, all total test scores are ranked and an upper $(U)$ group and a lower $(L)$ group is selected. The $U$ group contains the upper $27 \%$ of respondents and the $L$ group contains the lower $27 \%$ of respondents in this ranking. Taking the difference of the item difficulty index $P U$ for $U$ and $P L$ for $L$ results in the item discrimination index $D$ :

$$
\begin{equation*}
D=P U-P L . \tag{2}
\end{equation*}
$$

Results of this calculation are provided in Table I and range between 0.29 and 0.80 with a mean and standard deviation of $0.58 \pm 0.15$, respectively.

Similarly to the results of the internal consistency and item difficulty index, we find that the item discrimination indexes of the first eight items in the table are significantly smaller, indicating low discriminatory power for all TG and GT transition items.

The second way to describe the discriminatory power of the test but now only at the full test level is Ferguson's delta $\delta$. It represents the ratio of the interperson differences to the maximum number possible:

$$
\begin{equation*}
\delta=\frac{(k+1)\left(n^{2}-\sum_{i=1}^{k} f_{i}^{2}\right)}{k n^{2}}, \tag{3}
\end{equation*}
$$

with $k=48$ the number of items in the test, $n=385$ the sample size and $f$ the frequency of a certain score $i$, with $i$ from 1 to $k$. The resulting value for $\delta$ is 0.99 which easily satisfies the minimum required value of 0.9 , meaning the test offers good discriminatory power [40].

## B. Test results

## 1. Main effects and interaction effects

To analyze the data, we use the generalized estimating equations (GEE) approach [43,44], which is a specific variant of logistic regression analysis. This approach is required since the study design includes repeated observations within the series of responses of each respondent, for each variable that we study. The GEE method-in
contrast to regular logistic regression-includes adjustments for the correlations between the repeated dichotomous observations [45], which might otherwise result in incorrect conclusions. The analysis is performed with IBM ® SPSS ® Statistics version 24.0.0.0 software in which the binomial probability distribution and the logit link function are selected, which are the appropriate choices for our research design. Additionally, the parameter estimation method selected is Fisher; type III error is selected for analysis type; Wald statistics is chosen for the chisquare statistics; and the Bonferroni correction is selected to adjust for multiple comparisons. We include four factors in the model: representational transition, function type, context, and gender. Gender is included as a subject variable since Meltzer [11] has already shown there can be an interaction effect between gender and representation. The full model is analyzed, i.e., all factorial terms were included: main, 2 -way, 3 -way, and 4 -way effects.

Results show a significant main effect of the representation transition, with Wald $\chi^{2}(5)=437.063$ and $p<$ 0.001 . The pairwise comparisons indicate that transitions between tables and graphs are by far the easiest to perform in either direction. The mean accuracy for TG is $0.87 \pm$ 0.01 and for GT it is $0.86 \pm 0.01$. Transitions between formulas and tables are most difficult to perform with a mean accuracy for FT of $0.57 \pm 0.02$ and for TF of $0.56 \pm 0.02$. Finally, the difficulty of translating between formulas and graphs falls in between with a mean accuracy for FG of $0.64 \pm 0.02$ and for GF of $0.61 \pm 0.02$. There is a noticeable symmetry for representational transitions (e.g., GF and FG). This symmetry will be discussed in more detail later. Pairwise comparison between the transitions also supports this assessment since these comparisons always show a significant difference with $p<0.001$ with an exception between FT and GF which is $p=$ 0.002 -but none of the "symmetric pairs" differ significantly. Overall, accuracy is higher when a graph is involved and lower when a formula is involved; and translation to a formula is the most difficult direction. The analysis also shows a significant main effect of function type with Wald $\chi^{2}(3)=158.923$ and $p<0.001$. The pairwise comparison indicates that the two function types with a negative value for either $y$ intercept or slope-NP and PN-were most difficult with mean accuracies of $0.66 \pm 0.01$ and $0.66 \pm 0.02$, respectively. The function type with the highest mean accuracy is the directly proportional one with zero $y$ intercept and positive slope ( ZP ), with a value of $0.76 \pm 0.01$. A close second is the combination of positive $y$ intercept and positive slope (PP), which resulted in a mean accuracy of $0.74 \pm 0.01$. Overall, the use of negative signs for $y$ intercept or slope tends to result in significantly lower accuracies by about $10 \%$, and respondents achieve the highest accuracies for the directly proportional function type. Pairwise comparisons between the function types show that they all significantly differ

TABLE II. Mean accuracies for the interaction effect of representational transition (TG, GT, FT, TF, FG, and GF) and function type (NP, ZP, PP, and PN). The $95 \%$ confidence intervals for these numbers are always within $\pm 0.04$ and standard errors are between 0.013 and 0.021 . Students have greater difficulties with transitions including a formula ( F ) and/or function types with negative $y$ intercept or slope-NP and PN. There is a significant ( $p<0.01$ ) asymmetry for the ZP function type between FT and TF transitions.

|  | TG | GT | FT | TF | FG | GF |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| NP | 0.85 | 0.85 | 0.48 | 0.51 | 0.57 | 0.56 |
| ZP | 0.89 | 0.89 | 0.69 | 0.60 | 0.71 | 0.63 |
| PP | 0.86 | 0.84 | 0.62 | 0.63 | 0.71 | 0.71 |
| PN | 0.86 | 0.87 | 0.48 | 0.48 | 0.56 | 0.54 |

with $p<0.001$ except for the comparison between ZP and PP which do not differ significantly with $p=0.067$. Furthermore, a significant main effect of the context is found, with Wald $\chi^{2}(1)=5.831$ and $p=0.016$. The pairwise comparison indicates that the mean accuracy in mathematics is significantly ( $p=0.015$ ) higher than that in physics with respective accuracies of $0.72 \pm 0.02$ and $0.69 \pm 0.02$. Lastly, gender has no significant main effect with Wald $\chi^{2}(1)=1.960$ and $p=0.161$ for which the pairwise comparison indicates the mean accuracy for males is $0.72 \pm 0.02$ and for females $0.69 \pm 0.02$, which is not significantly different with $p=0.162$. The $95 \%$ confidence intervals on all of these values are within $\pm 0.03$, except for gender where the values are within $\pm 0.04$.

In addition to the main effects the GEE analysis also indicated four significant interaction effects. The most important one is the interaction between representational transition and function type with Wald $\chi^{2}(15)=137.672$ and $p<0.001$. These results are shown in Table II and in Fig. 3.


FIG. 3. Data from the 2-way interaction effect of representational transition (TG, GT, FT, TF, FG and GF) and function type (NP, ZP, PP and PN). The data points connected by lines represent the pairs of transitions which are each other's opposite. The $p$ value of the comparisons is indicated by $*$ for $p<0.05,{ }^{* *}$ for $p<0.01,{ }^{* * *}$ for $p<0.001$. There are two significant asymmetries for ZP function type. The first between FT and TF transition with $p<0.01$, and the second between FG and GF transitions with $p<0.05$.


FIG. 4. Data from the 3-way interaction effect of representational transition (TG, GT, FT, TF, FG and GF), function type (NP, ZP, PP, and PN) and context for the context of physics. The data points connected by lines represent the pairs of transitions which are each other's opposite. The $p$ value of the comparisons is indicated by * for $p<0.05$, ** for $p<0.01,{ }^{* * *}$ for $p<0.001$. There is a significant ( $p<0.05$ ) asymmetry for the ZP function type between FT and TF transitions in the physics context.

The table and the graph clearly show the dominant patterns of the main effects for representational transition and function type again. For transitions between tables and graphs all accuracies are very high and differences are very small. The more interesting part of the interaction between representational transition and function type is the significant asymmetry for the ZP function type between FT and TF transitions and between FG and GF transitions. There are two things to note here: first, the asymmetry is consistent in the way that the accuracy is always the lowest when transitioning to a formula. Second, it is unexpected that this effect is so large for the ZP function type, which is the function type with the highest mean accuracy. A closer look at this effect for physics and mathematics separately reveals that most of it is confined in mathematics as is shown in Figs. 4 and 5, which essentially show the threeway effect of representational transition, function type, and context for which Wald $\chi^{2}(15)=23.467$ and $p=0.075$.


FIG. 5. Data from the 3-way interaction effect of representational transition (TG, GT, FT, TF, FG, and GF), function type (NP, ZP, PP, and PN) and context for the context of mathematics. The data points connected by lines represent the pairs of transitions which are each other's opposite. The $p$ value of the comparisons is indicated by $*$ for $p<0.05, * *$ for $p<0.01, * * *$ for $p<0.001$. In the mathematics context there are three significant asymmetries: for the ZP function type there is one between FT and TF transitions with $p<0.001$ and one between FG and GF transitions with $p<0.001$; for NP there is a significant asymmetry between FT and TF with $p<0.05$.


FIG. 6. Data from the 2-way interaction effect of function type and gender. There are no significant differences between female and male for any function type.


FIG. 7. Data from the 3-way interaction effect of function type, context, and gender. There are no significant differences between physics and mathematics for any gender and function type combination.

Another significant interaction effect revealed by the analysis is the two-way interaction of function type and gender for which Wald $\chi^{2}(3)=10.089$ and $p=0.018$, of which the results are shown in Fig. 6. The pattern from the main effect for function type is followed for both genders separately. The largest differences between them are for the NP and PN function types, which are also the two function types with the lowest mean accuracy. There are no interesting significant pairwise comparisons of note for this interaction.

The next significant interaction effect is the three-way interaction of function type, context, and gender with Wald $\chi^{2}(3)=8.207$ and $p=0.042$, for which the results are shown in Fig. 7. There are no interesting significant pairwise comparisons of note for this interaction. The most interesting here is that for the NP and PN function types there are slightly larger differences for males between physics and mathematics, and also that these function types correspond with the largest differences between female and male scores. For ZP and PP function types the scores are very similar; of note is that the accuracy in physics is slightly higher than in mathematics in one single case: for males solving ZP items.

The last significant interaction effect is the 4-way interaction of all factors: representational transition, function type, context, and gender which has Wald $\chi^{2}(15)=$ 27.825 and $p=0.023$; this is shown in Fig. 8. The patterns of the main effects are still dominant here and the only significant difference between a representational transition and its opposite is with males in mathematics between the


FIG. 8. Data from the 4-way interaction effect of representational transition, function type, context and gender. (a) Mathematics*Female; (b) Mathematics*Male; (c) Physics*Female; (d) Physics*Male. The $p$ value of the comparisons is indicated by * for $p<0.05$, ** for $p<0.01,{ }^{* * *}$ for $p<0.001$. The only significant asymmetry is for the ZP function type between FG and GF transitions in the mathematics context with males.

TABLE III. Frequency (in percentage) of the chosen function type (NP, ZP, PP, PN, NN, or ZN) in the answer as function of the function type ( $\mathrm{NP}, \mathrm{ZP}, \mathrm{PP}$, and PN ) in the question from all 385 respondents. Correct answers are in bold face.

|  | NP | ZP | PP | PN | NN | ZN |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| NP | $\mathbf{6 4 . 0}$ | 5.2 | 6.1 | 9.5 | 9.5 | 5.7 |
| ZP | 3.5 | $\mathbf{7 3 . 7}$ | 8.4 | 4.4 | 3.5 | 6.5 |
| PP | 5.0 | 8.3 | $\mathbf{7 2 . 7}$ | 5.0 | 5.6 | 3.5 |
| PN | 9.9 | 6.0 | 9.3 | $\mathbf{6 3 . 4}$ | 6.1 | 5.4 |

FG and GF transitions. There are other asymmetries as well-although not significant ones-they are mostly found in mathematics and almost exclusively for the ZP function type. In physics the largest asymmetry between transitions is again for the ZP function type, specifically with females. Furthermore, the scores are spread out across a larger range for females as compared to males for both mathematics and physics.

## 2. Distractor analysis

Since the distractors for each question always contain the same six function types, the students' incorrect responses and their relation with representational transition and context can also be studied. Table III shows the frequencies for each of the distractors in the students' responses.

None of the distractors has a prevalence higher than $10.0 \%$, let alone higher than $16.7 \%$ which would be the prevalence if answers were chosen at random in a test with 6 distractors. It is nonetheless insightful to have a closer look at the most frequent errors. In Table III, three pairs of question and answer function types stand out. In the first (error E1) students choose NP in a PN question (9.9\%), and vice versa, choose PN in a NP question (9.5\%). This error can be explained by either a "sign switching" error, in which the students switch positive and negative for both the intercept and the slope; or it can be explained by "concept switching," in which students switch intercept and slope. Analysis at the student level reveals that just shy of $30 \%$ of students who make this error, do so symmetrically: switching NP for PN and also switching PN for NP in the same context and for the same representational transition. This indicates some persistence to making a symmetric E1 error.

The second pair (error E2), identifies those who choose PP in a PN question (9.3\%), and those who choose NN in an NP question ( $9.5 \%$ ), which can be explained by sign switching of the slope but at the same time correctly interpreting the sign of the intercept. Just over $20 \%$ of students who make such an error make it symmetrically:
they choose PP in a PN question and also NN in an NP question. This also indicates at least some persistence to making a symmetric E2 error, but less compared to the E1 error.

The last pair (error E3), shows switching of ZP and PP function types ( $8.3 \%$ and $8.4 \%$ ) which suggests difficulty distinguishing between zero or positive intercept with a positive slope. Analysis at the student level shows that very few students make a symmetrical E3 error, most often students make this error only once.

In summary, we have identified the following three errors as the predominant errors made in the test, but since the prevalences in the total population are very low, there is no indication of any fundamental issues with teaching this subject matter:
(i) (Symmetric) concept switching of intercept and slope.
(ii) (Symmetric) sign switching of the slope or of both slope and intercept.
(iii) Mostly nonsymmetric switching of ZP and PP function types.
More detailed data of the distractors as a function of context and representation transition are provided in Table IV in the Appendix in addition to a brief description of the most notable data points. Since these data are highly specific they are not discussed further in the main body of the paper.

## V. DISCUSSION

## A. Validation

The validation results of the test show that the test is consistent, reliable, has a good distribution of item difficulty, and good discriminatory power on a per item basis and for the whole test as well. This means this test is indeed appropriate to measure representational fluency for the specific use case of linear function types in physics and mathematics. In part, these good results are influenced by the relatively large number of questions (48) and their high similarity, which makes solving the test repetitive in nature. Furthermore, the very focused topic under study also helps in achieving these results. The TG and GT transitions stand out with their consistently high item difficulty indexes $P$ and low item discrimination indexes $D$, signaling that these items might be too easy to be included in the test. Since these items are part of the research questions, they are not omitted or replaced by other items.

## B. Results

All factors in the test design prove to have a significant main effect in the statistical model, thus supporting the inclusion of each one in the test design and their influence on the students' performance. One factor that is not in the test design but is included in the analysis, is gender; this
factor proves to be nonsignificant as a main effect but is included in three out of four significant interaction effects, each time accompanied by the function type factor. This supports the inclusion of gender as a factor, which was inspired by the work of Meltzer [11] and again confirmed by our results. Teachers should be aware of this difference. In practical terms this implies that females could benefit from extra support for the most difficult function types (NP and PN) and/or the most difficult context (physics). Our research design does not allow us to specify how such extra support might differ between females and males, though. Although we do not find a significant gender gap in our test, other studies have shown gender gaps to exist and persist in various concept inventories and tests with a broader scope than ours [46-49]. Recent studies [50,51] point to the use of a multidimensional context and the use of diagrams (phase diagrams, free body diagrams, multidimensional graphs, etc.) in the questions as factors which can create a bias in favor of male respondents, thus offering a possible explanation for the gender gap. Our results are complementary to these observations, since we find no significant gender gap in a test in which the questions are all restricted to a single dimension and in which no diagrams (other than 1D graphs) are present.

The main results are those concerned with representational fluency and the influence of function type and context. As was already clear from the validation, the TG and GT transitions are by far the ones with the highest accuracies, even so much that there is little distinction when comparing function types or contexts for these two transitions. Students are clearly very fluent at switching between tables and graphs-both ways-for all four function types in both contexts, which is in agreement with the results from De Bock, Van Dooren, and Verschaffel [15]. This can likely be explained by the direct visual correspondence between the numbers in the tables and the signs of the axes in the graphs which facilitates these transitions. In contrast, the transitions going from or to a formula are significantly more difficult for all function types and contexts, with the FT and TF transitions achieving the lowest mean accuracies, and FG and GF slightly higher ones. The cause of these lower accuracies is probably the higher abstraction level of formulas, which is more difficult to relate with numbers. Contrary to our results, De Bock, Van Dooren, and Verschaffel consistently found FG to be the single transition with the lowest mean accuracy for their selection of function types. A likely explanation for this difference is their inclusion of an inversely proportional function type and their exclusion of function types with negative $y$ intercept, since our results show that function types with negative values for either $y$ intercept or slope result in significantly lower mean accuracies, thus skewing these accuracies towards lower values, in particular for FT and TF transitions. An important difference between our test and the one from De Bock,


FIG. 9. Comparison of the data from De Bock, Van Dooren, and Verschaffel [15] (DB) and our data (Test) for the ZP function type for each of the representational transitions.

Van Dooren, and Verschaffel [15] is the abstract notation in the formula representation. We chose to make use of "greater than zero," "less than zero," and "equal to zero" instead of using numbers, which seems to be more difficult for students. Figures 9, 10, and 11 show the comparisons of the function types ZP, PN, and PP, which are investigated in each study. These figures show larger differences whenever formulas are involved, which suggests that our more abstract representation does negatively impact students' performance when compared to the use of numbers in formulas. Furthermore, they consistently show the largest differences between the two studies for the FT and TF transitions and smaller differences for FG and GF transitions. A plausible explanation for this is that the graphs mainly show the first quadrant which makes it easier to visually distinguish positive and negative slope, while in tables the use of negative numbers for the dependent variable (fourth quadrant) makes it more difficult to identify the sign of the rate of change. This in combination with the more abstract formulas might explain the observed differences. An additional issue with our more abstract notation is that the "less than" symbol might be more intuitive for students than the "greater than" symbol since we read and write from left to right and tend to sort numbers from smallest to largest. This means that students

PN


FIG. 10. Comparison of the data from De Bock, Van Dooren, and Verschaffel [15] (DB) and our data (Test) for the PN function type for each of the representational transitions.

PP


FIG. 11. Comparison of the data from De Bock, Van Dooren, and Verschaffel [15] (DB) and our data (Test) for the PP function type for each of the representational transitions.
might find $a<0$ and $0<a$ easier to interpret than $0>a$ and $a>0$, for example. This is only speculative though, and could be an interesting topic for future research. There are only two cases in which our results are higher than those from De Bock, Van Dooren, and Verschaffel: the TG and FG transitions for the PN function type. In the TG situation, the difference is just within our $95 \%$ confidence interval, but in the FG case the difference is more significant. The latter case appears to be a unique situation in which the use of numbers actually negatively affects students' performance. The explanation for this is possibly linked to the negative slope of the PN function type in combination with the lack of numbers in our FG items.

Ideally, representational fluency between two representations is "bidirectional," meaning that both directions of the transition should be equally fluent. In other words, the part "the ability to translate between representations" of our definition of representational fluency is ideally a symmetric ability. When studying this symmetry, we found that the accuracy for the directly proportional function type (ZP) has significant asymmetries since the transition to a formula has significantly lower accuracies when compared to the transition starting from a formula. This effect is present in both contexts, but is more pronounced in the mathematics context. This means that the function type with the highest mean accuracy (ZP) has the largest asymmetry-which is an unexpected result. In relation to the representational transitions, the asymmetry is very pronounced with the most difficult transitions, i.e., FT/TF and FG/GF transitions. This means that the pairs of representational transitions with the lowest mean accuracy-FT/TF and FG/GF-have the highest asymmetry and that the pair of representational transitions with the highest mean accuracy-TG and GT-have the lowest asymmetrywhich is more in line with what one would expect.

When comparing physics and mathematics our results indicate a significant difference of 0.03 between the mean accuracies for mathematics and physics, in favor of mathematics. This result is in line with that from Planinic et al. [8], who also found mathematics to be the easiest
context when compared to physics. Their study was for different concepts though: graph slope and area under a curve. Furthermore in another study [10] of theirs-in which they studied both positive and negative slope of line graphs in physics and mathematics-they also found that questions were solved better in a mathematics context compared to a physics context and that questions with negative slope were more often answered incorrectly. Our data do not really allow us to identify why students experience more difficulties with physics compared to mathematics, but comparing Figs. 4 and 5 does reveal that the largest discrepancies between the two contexts are with the NP questions with FT/TF transitions and also for the ZP questions with the FG transition.

Our findings provide accurate answers to research questions 1 and 2. The answer to research question 3 is that the most important significant interaction effect we find is the 2-way interaction between representational transition and function type; furthermore, there is a significant 2 -way interaction between function type and gender; a significant 3 -way interaction between function type, context, and gender; and a significant 4-way interaction between all factors involved: representation transition, function type, context, and gender.

## C. How can these findings be used in teaching?

Here, we link our theoretical findings to practice and teaching since it has often been acknowledged that there is a gap between educational research and practice [52].

The key idea is that the use of MERs can increase conceptual understanding [25]; and, specifically, the sequence of learning with MERs: the order in which a student learns to describe and comprehend the same concept in different representations. An optimal sequence would be scaffolded-since it could reduce the learner's cognitive load [53]-and start with the easiest representation and the easiest transitions and from there on increase in difficulty. This can be useful when the goal is to learn about the relation between representations-i.e., how elements in one representation relate to elements in another representation-and not necessarily to learn about the transition between them-i.e., the exact method or procedure to do the transition, which is more complex. This is in line with Ainsworth's second design principle or use for learning with MERs [24], i.e.: "automating translation when MERs are used to constrain interpretation." Possible methods are automatic translation or dynamic linking [25]-i.e., the process where the manipulations performed by a learner in one representation are automatically translated by a third party (such as a teacher or a computer) to changes in another representation thus freeing the learner to learn about the relation between the representations instead of focusing on the more difficult task of translation [ 54,55$]$. For example, a computer applet could be designed in which a student can manipulate a linear curve in a graph
by shifting it or changing the slope and immediately see the matching changes to a formulaic expression that represents the same linear relationship. Simple tasks such as trying to make one of the coefficients in the formula zero, positive, negative or a specific value can help students understand the relation between the representations and the meaning of certain elements in each representation. In the case of Ainsworth's third design principle or use for learning with MERs "scaffolding translation when MERs are used to develop deeper understanding," the stepwise increase in difficulty within a sequence would be complementary to the scaffolding of support by the teacher to gain deeper understanding. Thus, these design principles aim to reduce the cognitive load of the translation or aim to construct a carefully scaffolded increase in difficulty. For both purposes, a well-structured sequence can be of great practical use to teachers.

According [56] to McGee and Martinez-Planell [57], a good sequence is a "synergy of registers" [22] and not a "semiotic chain" [58]. In a synergy of registers one can perform the transitions between at least two different representations, which is called a "two-register synergy." Preferably more registers are included such as with the "three-register synergy" shown in Fig. 12(a). In a semiotic chain-Fig. 12(b), one only learns to perform the transitions along the chain which might result in lower understanding. Thus it is advisable [57] to explicitly add missing transitions in a semiotic chain. Not every transition must be included though; sometimes a "base representation" exists. This is a representation that is used as a layover representation in the transition process. In Fig. 12(c) "Formula" is the base representation and serves as a layover between, e.g., "Diagram" and "Table." Heckler and Scaife [59] studied vector addition and subtraction in the arrow and $i j k$ representation and presented empirical evidence for the existence of a natural hierarchy for teaching these operations in which the $i j k$ format can be a useful base representation. A base representation can be present during the learning process and can remain so thereafter, but ideally an "all-register synergy" in which all transitions are present is achieved. In short, it is best to have as many "closed loops" as possible in a sequence diagram.


FIG. 12. Hypothetical examples of (a) an all-register synergy; (b) a semiotic chain; (c) a synergy of registers with missing transitions and a central base representation (Formula). A formula can be any algebraic expression, and a diagram could be a free body diagram or a phase diagram, for example.

The average difficulties for the transitions in our data offer similar information as an all-register sequence diagram for tables, graphs, and formulas and suggest to start from graphs and tables when introducing linear relationships. In physics, one could think of plotting the $x(t)$ graph of a uniform linear motion using a motion sensor. The experiment will show a straight line graph that can immediately be coupled to the corresponding table with the measurement points. Using different initial positions and or velocities, all different function types for linear relationships can then be identified (NN, NP, ZN, etc.). Next, the formula representation can be introduced by transitioning from graphs and tables to formulas by first matching the shape of the curve with the formula; and, finally, the transitions between tables and formulas are treated. Furthermore, the general similarities for the representational transitions between the physics data and the mathematics data suggest that it would be best to do this in both contexts simultaneously. Ideally, all possible representations are then compared together at the same time so students can recognize the same function type in each representation. Such a summary exercise could be a perfect fit for the dynamic linking approach by creating a computer application which simulates or measures the motion of a moving object and represents the position and time data in various representations simultaneously. An important note here is the issue of acceleration when doing an experiment. A toy car starting from a standstill must first accelerate before it can achieve a nonzero constant velocity. This nonlinear behavior will show up in the measurement data. When doing such experiments it is highly recommended to use these phenomena as learning opportunities for students and teach them about acceleration and the limitations of a linear model when applied to reality. A useful comparison to make when doing such experiments is comparing a "rolling start" to a start from a standstill and see how the tables and graphs differ.

## D. Future research

In this section, we briefly discuss some possible future research directions as continuations and refinements of this study. First, the test structure can be easily applied to other representations, other transitions, other function types-or a different topic altogether-and other contexts. More subtle modifications to the current test can also be made, such as using different quadrants in the graphs and tables or including the roots in the graphs and tables, which might open up new errors such as mistaking the root for the $y$ intercept. Second, our study is more focused on the quantitative side, it would be insightful to complement this with qualitative research. A good suggestion would be thinking aloud interviews, which might provide a better understanding of the most frequently made errors. Third, the way we interpret the representational fluency in this multiple-choice test needs some caution. In each item, we
show respondents a specific function type in a certain representation, such as a formula. At the same time and on the same page or screen the respondents are presented with six alternatives in another representation, such as tables. We assume that this example studies the representational fluency of transitioning FT from a formula to a table but this might not be the way students actually solve the question. A student might identify the function type in the formula to a base representation (for example, a graph), thus performing a FG transition and then transition each answer option to a graph to find the matching graph, thus performing the TG transition a maximum of six times. This example might actually be very realistic since our research—and that of other cited in this work-shows that graphs are the easiest representation of the three under study, which might indicate that it serves as a base representation when transitioning between other representations. Alternatively - and following the same example-a student might transition all six answer options to formulas, thus performing the TF transition 6 times and not the FT transition as intended. Or lastly, a student might make the FT transition once and then find the matching table from the six options, which is how we interpreted the results. Other constructs are of course also possible. To understand the exact process in a student's reasoning, the proposed thinking aloud interviews could provide more insight into this issue. This also reconnects with the remark made in the introduction and literature section which was that instead of using the term representational fluency, the term representational recognition might also be appropriate since students might actually recognize the identified mathematical concept of the question in the answer possibilities instead of performing an active conversion of the identified mathematical concept.

## VI. CONCLUSION

This research resulted in a validated 48-item multiplechoice test to study representational fluency in physics and mathematics on the topic of linear functions. The test has a high Cronbach's $\alpha$ of 0.946 , a good distribution of item difficulty indexes and item discrimination indexes, as well as a very good Ferguson's delta of 0.99 . Our results show significant main effects for representational transition, function type, and context. Students prove very adept in transitions between tables and graphs; they have the most difficulty with transitions which included a formula. Whenever a function type has a negative sign in either the $y$ intercept or the slope, the mean accuracy was significantly lower, which indicates that students have significantly more difficulties with negative numbers. When comparing contexts the analysis results in a significant difference between physics and mathematics. The GEE analysis also shows four significant interaction effects, of which the most important one is between representational transition and function type. The other significant interaction effects
always include gender and function type. The main trend there is that the difference between male and female accuracies is larger for the most difficult function types and for the most difficult transitions. When comparing the symmetry of the representational transition between any two representations, the analysis clearly shows a significant asymmetry for the directly proportional function type. This unexpected result is most evident in the transitions between formulas and tables and between formulas and graphs in which the transition to a formula always has the lowest accuracy. Interestingly, the largest asymmetry is found for the-on average-easiest function type (ZP) and for theon average-most difficult representational transitions, which are the ones containing a formula. Furthermore, this effect is most pronounced in the context of mathematics. The errors students make most in these particular cases are selecting the function type with positive $y$ intercept and positive slope instead of zero $y$ intercept and positive slope. The distractor analysis results in the identification of three distinct dominant errors: (i) (symmetric) concept switching of intercept and slope; (ii) (symmetric) sign switching of the slope or of both slope and intercept; (iii) (mostly nonsymmetric) switching of directly proportional function type and the function type with positive $y$ intercept and positive slope. Overall, the analysis shows that students' knowledge is strongly dependent on representational transition, function type, and interactions. Furthermore, our study confirms similar results about the difficulty of various representational transitions from De Bock, Van Dooren, and Verschaffel [15] and highlights the additional difficulties students experience with our more abstract representation in formulas compared to the use of numbers in formulas in their study.

As a final note, we emphasize that this study focuses on a specific definition of representational fluency and the ability to transition between representations. Our results provide quantitative insight about this, but do not allow us to determine how students perform the transition. To better understand these processes, additional and more detailed information on students' thinking and their use of other representations is needed.

## ACKNOWLEDGMENTS

This research was funded by the IWT-SBO project STEM@school. For his help and insightful discussion we would like to thank Wim Van Dooren from KU Leuven.

## APPENDIX: DETAILED DISTRACTOR ANALYSIS

Table IV provides a highly detailed overview of the distractor data as a function of the context and the representation transition. The table clearly shows that most of the erroneous selections made by students are in the

TABLE IV. Frequency (in percentages) of the chosen function type in the answer (A) as function of the function type and the representational transition in the question (Q) from all 385 respondents for all questions with a physics context (left) and with a mathematics context (right). Correct answers are in bold face.

| Physics |  |  |  |  |  |  |  | Mathematics |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Q | A | TG | GT | FT | TF | FG | GF | Q | A | TG | GT | FT | TF | FG | GF |
| NP | NP | 84.7 | 84.9 | 44.7 | 44.7 | 56.1 | 53.5 | NP | NP | 86.2 | 84.9 | 52.5 | 58.7 | 58.4 | 59.2 |
|  | ZP | 2.3 | 2.6 | 11.4 | 6.8 | 12.2 | 3.6 |  | ZP | 1.6 | 2.3 | 3.9 | 4.7 | 6.0 | 4.4 |
|  | PP | 2.2 | 1.8 | 8.1 | 9.1 | 7.3 | 7.5 |  | PP | 0.8 | 2.3 | 11.4 | 8.3 | 8.3 | 6.0 |
|  | PN | 2.6 | 2.3 | 14.0 | 15.8 | 14.5 | 14.0 |  | PN | 2.3 | 2.9 | 11.2 | 11.9 | 13.2 | 9.6 |
|  | NN | 4.4 | 4.9 | 11.9 | 15.3 | 3.9 | 16.9 |  | NN | 5.7 | 4.7 | 11.7 | 11.9 | 6.5 | 15.8 |
|  | ZN | 3.9 | 3.4 | 9.9 | 8.3 | 6.0 | 4.4 |  | ZN | 3.4 | 2.9 | 9.4 | 4.4 | 7.5 | 4.9 |
| ZP | NP | 1.8 | 1.8 | 5.7 | 5.5 | 4.2 | 4.2 | ZP | NP | 1.0 | 1.3 | 3.4 | 4.4 | 4.2 | 4.7 |
|  | ZP | 90.4 | 89.6 | 67.0 | 61.6 | 67.3 | 63.1 |  | ZP | 87.5 | 88.3 | 71.7 | 59.2 | 75.1 | 63.6 |
|  | PP | 3.1 | 2.1 | 7.3 | 12.5 | 10.6 | 11.2 |  | PP | 4.9 | 5.7 | 8.3 | 18.4 | 7.0 | 9.6 |
|  | PN | 1.8 | 2.9 | 6.8 | 4.4 | 4.7 | 4.4 |  | PN | 2.6 | 1.3 | 8.3 | 4.9 | 5.2 | 5.5 |
|  | NN | 1.6 | 1.0 | 6.0 | 6.0 | 2.1 | 4.9 |  | NN | 2.3 | 1.8 | 1.8 | 6.0 | 2.9 | 5.7 |
|  | ZN | 1.3 | 2.6 | 7.3 | 10.1 | 11.2 | 12.2 |  | ZN | 1.6 | 1.6 | 6.5 | 7.0 | 5.7 | 10.9 |
| PP | NP | 1.0 | 2.1 | 4.7 | 8.8 | 6.5 | 7.0 | PP | NP | 2.3 | 2.1 | 5.5 | 7.0 | 6.2 | 6.5 |
|  | ZP | 6.0 | 6.0 | 18.7 | 5.7 | 10.1 | 5.5 |  | ZP | 4.4 | 8.8 | 14.3 | 6.8 | 7.8 | 5.2 |
|  | PP | 86.2 | 85.2 | 59.7 | 60.3 | 69.1 | 68.1 |  | PP | 84.9 | 82.3 | 65.2 | 65.5 | 73.2 | 72.7 |
|  | PN | 2.1 | 2.6 | 8.1 | 9.6 | 6.2 | 5.2 |  | PN | 3.4 | 3.4 | 6.5 | 4.4 | 3.9 | 4.2 |
|  | NN | 2.3 | 1.8 | 4.4 | 10.1 | 5.7 | 9.1 |  | NN | 2.1 | 2.3 | 4.7 | 11.2 | 6.0 | 7.5 |
|  | ZN | 2.3 | 2.3 | 4.4 | 5.5 | 2.3 | 5.2 |  | ZN | 2.9 | 1.0 | 3.9 | 5.2 | 2.9 | 3.9 |
| PN | NP | 3.4 | 3.4 | 12.5 | 10.6 | 14.5 | 11.4 | PN | NP | 1.6 | 2.1 | 13.8 | 15.1 | 14.3 | 15.8 |
|  | ZP | 3.1 | 3.9 | 10.1 | 9.1 | 10.4 | 5.2 |  | ZP | 3.6 | 1.6 | 8.1 | 6.8 | 7.0 | 3.6 |
|  | PP | 3.6 | 1.6 | 9.1 | 20.3 | 10.1 | 15.1 |  | PP | 5.2 | 3.9 | 9.6 | 14.0 | 7.8 | 11.4 |
|  | PN | 85.5 | 85.5 | 45.5 | 46.0 | 51.4 | 53.2 |  | PN | 86.2 | 88.3 | 50.9 | 51.4 | 60.5 | 55.8 |
|  | NN | 2.1 | 3.1 | 11.4 | 6.2 | 6.5 | 9.1 |  | NN | 0.8 | 0.5 | 10.4 | 8.1 | 5.5 | 9.1 |
|  | ZN | 2.3 | 2.6 | 11.4 | 7.8 | 7.0 | 6.0 |  | ZN | 2.6 | 3.6 | 7.3 | 4.7 | 4.9 | 4.2 |

physics context and mainly with the transitions which include a formula as was already clear from the validation and the discussion in the body of the paper. In this Appendix we will highlight the most notable data points in Table IV.

The highest frequency of an incorrect answer is in physics, where $20.3 \%$ of students chose PP in a PN question with a TF transition-whereas only $14 \%$ chose PP in the isomorphic question in mathematics. This means that the single most frequent error students made was when they had to identify a negative velocity in a table in physics and chose a positive velocity instead. This issue is likely explained by either the tendency to identify velocity as speed in Flemish-thus ignoring the minus sign—and/or the difficulty students experience with identifying a series of increasing numbers with negative signs. Since this erroneous choice was made substantially more often in physics, the issue with velocity and speed likely contributed to this difference between the contexts.

The second highest frequency in the table is that $18.7 \%$ of students chose ZP in a PP question in physics with the FT transition-for the isomorphic mathematical question this is $14.3 \%$. This means that these students, when confronted with a formula with all positive coefficients, identify this as a directly proportional function ZP.

The third highest number is $18.4 \%$ in ZP-TF questions in mathematics for which students chose PP-in physics the frequency is $12.5 \%$. These students-when confronted with a table including the point $(0,0)$-still chose a formula with a positive $y$ intercept. Comparing this with the frequency of only $4.9 \%$ under the TG transition, the difficulty here is most likely not from interpreting the table, but from translating to a formula. A possible explanation is that students might consider a ZP table as the stereotypical table for a linear function and at the same time consider the PP formula as the stereotypical formula for a linear function, which could cause a mismatch.
[1] A. R. P. de Atade and I. M. Greca, Epistemic views of the relationship between physics and mathematics: Its influence on the approach of undergraduate students to problem solving, Sci. Educ. 22, 1405 (2013).
[2] T. H. Kjeldsen and J. Lützen, Interactions between mathematics and physics: The history of the concept of function-teaching with and about nature of mathematics, Sci. Educ. 24, 543 (2015).
[3] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, Sci. Educ. 24, 561 (2015).
[4] E.F. Redish, Problem solving and the use of math in physics courses, in World View on Physics Education in 2005, Focusing on Change, Delhi (2005), pp. 1-10.
[5] H. T. Hudson and W. R. McIntire, Correlation between mathematical skills and success in physics, Am. J. Phys. 45, 470 (1977).
[6] E. B. Pollock, J. R. Thompson, and D. B. Mountcastle, Student Understanding Of The Physics And Mathematics Of Process Variables In P-V Diagrams, in Physics Education Research, American Institute of Physics Conference Series, Vol. 951, edited by L. Hsu, C. Henderson, and L. McCullough (AIP, New York, 2007), pp. 168-171, DOI: 10.1063/1.2820924.
[7] 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).
[8] 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).
[9] 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).
[10] 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).
[11] D. E. Meltzer, Relation between students' problem-solving performance and representational format, Am. J. Phys. 73, 463 (2005).
[12] 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).
[13] 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).
[14] M. De Cock, Representation use and strategy choice in physics problem solving, Phys. Rev. ST Phys. Educ. Res. 8, 020117 (2012).
[15] 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).
[16] E. Deliyianni, A. Gagatsis, I. Elia, and A. Panaoura, Representational flexibility and problem-solving ability
in fraction and decimal number addition: A structural model, Int. J. Sci. Math. Educ. 14, 397 (2016).
[17] 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).
[18] 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 Math. Educ. 41, 627 (2009).
[19] S. E. Ainsworth, P. A. Bibby, and D. J. Wood, Analysing the costs and benefits of multi-representational learning environments, Learning with Multiple Representations (Pergamon Amsterdam, 1998), pp. 120-134.
[20] R. Even, Factors involved in linking representations of functions, J. Math. Behav. 17, 105 (1998), representations and the Psychology of Mathematics Education: Part I.
[21] S. Ainsworth, P. Bibby, and D. Wood, Examining the effects of different multiple representational systems in learning primary mathematics, J. Learn. Sci. 11, 25 (2002).
[22] R. Duval, A cognitive analysis of problems of comprehension in a learning of mathematics, Educ. Studies Math. 61, 103 (2006).
[23] D. Kirsh, Thinking with external representations, AI \& SOCIETY 25, 441 (2010).
[24] S. Ainsworth, The functions of multiple representations, Comput. Educ. 33, 131 (1999).
[25] S. Ainsworth, Deft: A conceptual framework for considering learning with multiple representations, Learning Instr. 16, 183 (2006).
[26] D. Niemi, Assessing conceptual understanding in mathematics: Representations, problem solutions, justifications, and explanations, J. Educ. Res. 89, 351 (1996).
[27] 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).
[28] 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).
[29] J. D. Davis, Real-world contexts, multiple representations, student-invented terminology, and y intercept, Math. Think. Learn. 9, 387 (2007).
[30] A. H. Schoenfeld, J. P. Smith, and A. Arcavi, Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain, Adv. Instruct. Psychol. 4, 55 (1993).
[31] 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).
$[32]$ 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).
[33] 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).
[34] F. M. Goldberg and J. H. Anderson, Student difficulties with graphical representations of negative values of velocity, Phys. Teach. 27, 254 (1989).
[35] R. J. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. 62, 750 (1994).
[36] 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).
[37] See Supplemental Material at http://link.aps.org/ supplemental/10.1103/PhysRevPhysEducRes.14.020105 for the full physics and mathematics test translated to English.
[38] A. A. Nistal, W. Van Dooren, and L. Verschaffel, Improving students' representational flexibility in linear-function problems: an intervention, Educ. Psychol. 34, 763 (2014).
[39] An older version of this test had been administered to 148 students in a pilot study in 2016 as a first validation which resulted in good values for all relevant validation quantities. Because of some changes made to the test for a second administration in 2017, and due to the larger number of respondents in this second administration, the updated test is validated again with the new data from 2017.
[40] 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).
[41] A. C. Ornstein, Norm-referenced and criterion- referenced tests: An overview, NASSP Bull. 77, 28 (1993).
[42] R. J. McCowan and S. C. McCowan, Item analysis for criterion-referenced tests, Online Submission (ERIC, 1999), https://eric.ed.gov/?id=ED501716.
[43] K.-Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73, 13 (1986).
[44] M. Stokes, C. Davis, and G. Koch, Categorical Data Analysis Using the Sas®System, 2nd ed. (SAS Institute Inc., Cary, NC, USA, 2000).
[45] C.-f. Sheu, Regression analysis of correlated binary outcomes, Behavior Res. Methods, Instrum. Comput. 32, 269 (2000).
[46] J. Docktor and K. Heller, Gender differences in both force concept inventory and introductory physics performance, AIP Conf. Proc. 1064, 15 (2008).
[47] S. Andersson and A. Johansson, Gender gap or program gap? students' negotiations of study practice in a course in
electromagnetism, Phys. Rev. Phys. Educ. Res. 12, 020112 (2016).
[48] S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch, and N. R. Walet, Gender differences in conceptual understanding of newtonian mechanics: a uk cross-institution comparison, Eur. J. Phys. 34, 421 (2013).
[49] S. J. Pollock, Comparing student learning with multiple research based conceptual surveys: Csem and bema., AIP Conf. Proc. 1064, 171 (2008).
[50] K. Wilson, D. Low, M. Verdon, and A. Verdon, Differences in gender performance on competitive physics selection tests, Phys. Rev. Phys. Educ. Res. 12, 020111 (2016).
[51] H. Dawkins, H. Hedgeland, and S. Jordan, Impact of scaffolding and question structure on the gender gap, Phys. Rev. Phys. Educ. Res. 13, 020117 (2017).
[52] H. Broekkamp and B. van Hout-Wolters, The gap between educational research and practice: A literature review, symposium, and questionnaire, Educational Res. Eval. 13, 203 (2007).
[53] J. Sweller, J. J. G. Van Merrienboer, and F. G. W. C. Paas, Cognitive architecture and instructional design, Educ. Psychol. Rev. 10, 251 (1998).
[54] J. J. Kaput, Technology and mathematics education, in Handbook of Research on Mathematics Teaching and Learning, edited by D. A. Grouws (Macmillan Publishing Company, New York, NY, 1992), p. 515.
[55] M. Scaife and Y. Rogers, External cognition: how do graphical representations work?, International Journal of Human-Computer Studies 45, 185 (1996).
[56] They-and some other authors-describe an external representation as a semiotic register, the two terms can be used interchangeably.
[57] D. L. McGee and R. Martinez-Planell, A study of semiotic registers in the development of the definite integral of functions of two and three variables, Int. J. Sci. Math. Educ. 12, 883 (2014).
[58] N. Presmeg, Semiotics and the "connections" standard: Significance of semiotics for teachers of mathematics, Educ. Studies Math. 61, 163 (2006).
[59] A.F. Heckler and T. M. Scaife, Adding and subtracting vectors: The problem with the arrow representation, Phys. Rev. ST Phys. Educ. Res. 11, 010101 (2015).


[^0]:    *stijn.ceuppens@kuleuven.be
    'mieke.decock@kuleuven.be

[^1]:    ${ }^{1}$ An older version of this test had been administered to 148 students in a pilot study in 2016 as a first validation which resulted in good values for all relevant validation quantities. Due to some changes made to the test for a second administration in 2017, and due to the larger number of respondents in this second administration, the updated test is validated again with the new data from 2017.

