# Testing students ability to use derivatives, integrals, and vectors in a purely mathematical context and in a physical context

Marta Carli[1,†] Stefania Lippiello,[2] Ornella Pantano,[1]
Mario Perona,[3] and Giuseppe Tormen[1,*]

[1]*University of Padua, Department of Physics and Astronomy, Via Marzolo 8, 35131 Padua, Italy*
[2]*Liceo Scientifico J. Da Ponte, Via S. Tommaso d'Aquino 12, 36061 Bassano del Grappa (VI), Italy*
[3]*Istituto di Istruzione Superiore C. Marchesi, Viale Codalunga 1, 35138 Padua, Italy*

In this article, we discuss the development and the administration of a multiple-choice test, which we named *Test of Calculus and Vectors in Mathematics and Physics* (TCV-MP), aimed at comparing students' ability to answer questions on derivatives, integrals, and vectors in a purely mathematical context and in the context of physics. The comparison between the two contexts was achieved by using parallel (isomorphic) questions in mathematics and physics. The final version of the test contains 34 items (17 in a purely mathematical context and 17 in the context of physics) involving different representations (graphs, words, numbers, and formal expressions) of the concepts covered by the test. The test was administered in Spring 2018 to 1252 first-year students enrolled in 23 different degree programs of the School of Science and the School of Engineering of the University of Padua. We assessed the validity, reliability, and discriminatory power of the test both as a whole and at the single-item level, obtaining values within the desired ranges. The analysis of students' answers to individual items and the comparison between parallel mathematics and physics items provides insights into the factors that affect students' ability to use derivatives, integrals, and vectors in the context of introductory physics. We believe that the instrument we have developed can be useful not only for research purposes, but also for instructors and for students.

## I. INTRODUCTION

Physics relies heavily on mathematics for both describing and making predictions about phenomena. Many views of the role of mathematics in physics can be listed, such as a pragmatic tool, a language [1], and a way of reasoning [2].

Although mathematics is undoubtedly a powerful tool for calculation in physics, its role cannot be reduced to technical aspects. Ever since modern science was born, mathematics and physics have been strongly intertwined, so that the role of mathematics in physics can best be described as structural [3–5]. Despite that, in most universities mathematics is seen just as a prerequisite for physics [6]. With only a few exceptions [7–9], calculus and physics are typically treated as separate subjects taught by faculty belonging to different departments, and physics instructors expect calculus courses to provide the students with the mathematical tools they need in physics. However, it is well known that proficiency in math does not guarantee success in physics [1,10–15]. Even when students complete a calculus course successfully, they may have difficulties in using the same mathematical tools in a physical context, and they can be proficient in the "technical" use of mathematics without actually making sense of physics [5,16,17].

Students' ability to use mathematics in the context of physics has often been framed in terms of "transfer" [18–20]. Traditionally, transfer has been defined as the ability to apply previously acquired knowledge in a new context. However, this definition does not account for all the phenomena associated with knowledge building and the current theories of knowledge [21,22]. Rather than a mere application of previous knowledge, modern perspectives view transfer as an active, student-centered dynamic process, governed by the students' epistemic frames [5] and "noticing" of relevant problem features [23,24], the visual attributes and "affordances" of the problem [25–28], and dependent on the disciplinary context [29]. For example, in the preparation for future learning (PFL) perspective [30] evidence of transfer is not sought in "one-shot" students' performances, but rather in the whole process of learning. In the transfer in pieces (TiP) perspective [31], more subtle evidence of transfer are acknowledged, so that even incorrect answers, when looked through a more fine-grained lens, can bring the signs of transfer.

*Deceased.
†marta.carli.1@unipd.it

Finally, the actor-oriented transfer perspective [32] focuses on identifying what similarities, if any, a student sees or dynamically constructs in two connected but different situations. These different perspectives are not mutually exclusive: multiple perspectives can be combined and used to highlight different aspects of the problem [22,33–36].

In recent years, "conceptual blending" has been proposed as an alternative framework to account for both context dependency and the relevance of prior knowledge in problem solving, both in physics and in other branches of science [2,37–40]. According to this framework, a student who encounters a problem in a new context "blends" information from different "mental spaces" (e.g., mathematics, physics, everyday experience) to construct an emergent "blended space" that is unique to the problem and that is used to solve it. In this view, many physical concepts are inherently "blended objects" in that they are best represented by a joint mathematical-physical model that can be described at different degrees of mathematization [3].

In this complex background, we may therefore wonder to what extent students' difficulties in the mathematization of physics are due to a lack of understanding mathematics. Research has tried to find an answer to this question tackling the issue on multiple fronts. In this work, we contribute to this effort by proposing a quantitative instrument, a test which we named the test of calculus and vectors in mathematics and physics (TCV-MP), aimed at highlighting differences in students' answers to isomorphic mathematics and physics questions involving mathematical concepts that are typically encountered in introductory physics: derivatives, integrals, and vectors.

## II. THEORETICAL BACKGROUND

Several accounts in the literature on physics education report that students have difficulties in understanding the relationship between kinematical quantities, a problem that becomes particularly relevant when graphs are involved [41–44].

While a general categorization of students' difficulties with graphs was proposed by Leinhardt *et al.* [45], the first taxonomy of these difficulties in the context of kinematics was proposed by McDermott and colleagues [46]. A decade later, Beichner [47] designed the test of understanding graphs in kinematics (TUG-K), a multiple-choice test based on six dimensions corresponding to different categories of typical students' mistakes that he identified: graph as picture errors, confusion between slope and height, variable confusion, non-origin slope errors, area ignorance, and confusion among area, slope, and height. Recently, Zavala *et al.* [48] have proposed a modified version of the TUG-K improving the parallelism between the different dimensions of the test, and Dominguez *et al.* [49] have developed the test of understanding graphs in calculus (TUG-C), the counterpart of the TUG-K in a purely mathematical context.

The quantitative studies mentioned above were aimed at identifying taxonomies of students' mistakes, useful for a broad characterization of students' difficulties. Other studies sought to investigate specific topics in a greater detail, making use of qualitative methods in order to gain insights into students' reasoning. For example, Wemyss and Van Kampen [50] investigated students' ability to determine the direction of motion, the constancy of speed, and a numerical value for the speed of an object at a point on a numerical linear distance-time graph. They found that "technical" mathematical issues could not account for all of the observed difficulties, and that the incorrect prior learning in physics also played a crucial role. Bollen *et al.* [51] confirmed Wemyss and Van Kampen's results, and pointed out that having a qualitative understanding of a distance-time graph is not sufficient to correctly determine a value for the speed. Concerning integral-related concepts, Nguyen and Rebello [52] researched students' difficulties in using the concept of area under a curve in physics problems. Even when the students mentioned the concept, they were not always able to relate it to the process of accumulation. Similar results had been found previously in the context of mathematics [53].

Another research strand is focussed on comparing students' ability to solve problems in different contexts. For example, Christensen and Thompson [54] investigated students' understanding of slopes and derivatives using graphical "physics-less physics questions", i.e., physics problems stripped of their physical context. They found that students had difficulties in conceptualizing mathematics tasks formulated in this way, and they argued that "the type of mathematical tasks we want our students to do in a physics class may simply be foreign to their mathematical ways of thinking". Jones [55–57] compared students' problem-solving strategies in problems involving definite integrals in mathematics and in physics. He found that students, in general, rely more often on antiderivative or area-based ideas rather than on Riemann sum-based conceptions; in mathematics this was not a problem, since the three conceptualizations were equally effective, whereas in physics Riemann sum-based ideas were more productive, but underutilized. He also found that students often hold a "prototype image" of integrals that typically considers only positive values, with little variation in both the size of function values and steepness of the graph, and does not include special cases such as discontinuities. The author argues that this fact can lead to difficulties in interpreting integrals when the given curve differs from this prototype [58,59]. Finally, researchers at the University of Zagreb [29,60–63] explored students' strategies in interpreting and using graphs in mathematics, in physics, and in a context other than physics, using sets of isomorphic items on the concepts of slope and of area under a curve. They found that students' strategies for interpreting the graphs were context dependent and domain specific. Interestingly,

students were able to use a variety of productive strategies in mathematics and in contexts other than physics, while in physics they tended to stick to previously learned strategies and to use formulas.

Another concept students typically struggle with is the vectorial nature of many physical quantities. Knight [64] found that students had difficulty in manipulating vector components, establishing the direction of a vector, dealing with vectorial sum, and using the different types of vector products. More recently, Nguyen and Meltzer [65] confirmed that, even after explicit instruction on vectors, many students have conceptual confusion in vector concepts (vector sum, magnitude, and direction), particularly when vectors are represented in graphical form.

Also in the case of vectors, some studies have attempted to provide taxonomies of students' difficulties using quantitative instruments. One such taxonomy was proposed by Barniol and Zavala [66], who developed the test of understanding of vectors (TUV). Their taxonomy includes the graphical properties of direction, magnitude, and components of a vector (e.g., components of unit vectors); the graphical procedures of vector operations (e.g., difference between vectors, multiplication by a negative scalar); calculations that involve angles, trigonometric functions, and the Pythagorean theorem (e.g., confusion between sine and cosine); and calculations of dot and cross products that involve unit-vector notation. Recently, Susac and colleagues [67] have administered the TUV to 889 first-year students and they have found that some of the items were very difficult for the students: the most difficult vector concept was the unit vector, followed by the cross product, subtraction of vectors, the dot product, and vector direction.

Concerning the use of vectors in the context of physics, Flores and colleagues [68] found that, even after instruction in mechanics, many students were unable to determine the direction of the difference between two velocity vectors, to find the direction of the acceleration vector, and to determine the relationship between individual forces and the net force acting on an object. Shaffer and McDermott [69] also investigated students' ability to treat velocity and acceleration as vectors and they found several difficulties, including not recognizing that the velocity vector is tangent to the trajectory, not distinguishing between velocity and acceleration, assuming that the acceleration is zero because the speed is constant, using a nonzero vector for the velocity and a zero vector for acceleration at a turnaround point, and not associating the direction of acceleration with the direction of net force.

In order to compare students' ability to use vectors in a purely mathematical context and in the context of physics, Van Deventer and colleagues [70,71] designed isomorphic mathematics and physics items for a subset of vector-related concepts (vector magnitude and components, vector subtraction, dot product). Students' overall performance was similar in the two contexts, but some specific differences

were observed. For instance, students performed better on vector subtraction in physics than in mathematics, since the reference to a coordinate system cued them towards a correct answer; conversely, when asked to find an algebraic expression for the $x$ component of a vector, students struggled more in physics than in mathematics when the given angle was between the vector and the $y$ axis. Finally, students performed poorly on questions involving the dot product both in mathematics and in physics, with many students drawing a vector for the dot product.

For both calculus and vectors, the representational format (graphs, words, numbers, equations) in which a problem is formulated seems to be particularly relevant. Representational fluency (i.e., the ability to dynamically pass from one representational format of a concept to another) is very important in physics and it is considered a sign of expertise [72,73]. Specific instruments for assessing students' representational fluency have been designed, such as the KiRC inventory for kinematics [74], and, on a more interdisciplinary perspective, the representational fluency survey [75].

Research is unanimous in highlighting that passing from one representation to the other is not straightforward for students, and that they may adopt different problem-solving strategies depending on the specific representational format used in the problem [76–78]. In fact, some authors have compared students' problem solving abilities between variants of the same test item formulated using different representational formats, and they have observed statistically significant differences [79–82].

In particular, researchers have investigated the differences between students' ability to handle formal representations and their ability to use graphs: these two representational formats are particularly relevant for physics, and among the different possible representations, they are the ones that involve the highest degree of mathematization [3,83]. For example, Bajracharya [78] compared students' use of integral-related concepts in mathematics and in physics, in the presence of graphs and/or equations. He found that, in general, students preferred to solve problems analytically rather than using graphical reasoning, and even when no analytic information was included in a problem, some students still attempted to solve it by inferring an analytic expression from the graph. In addition, some of the students who tried to use the graphs to solve the problem either used irrelevant features or tried to read off numbers directly from the graph rather than engaging in interpretation of graphical properties. Finally, using eye tracking the author found that previously reported common mistakes could be cued by specific problems features such as the presence or absence of equations or the notation used in the problem. More recently, Van den Eynde and colleagues [84] investigated students' ability to translate between graphs and equations both in a purely mathematical context and in the context of physics. They found that the students had fewer difficulties on

mathematics items than on physics items, and that they performed better on items starting from a graph than in those starting from an equation.

Concerning the topic of vectors, Heckler and Scaife [85] explored how students' understanding of vectors in mathematics and physics could be influenced by the representation used in the problem. They found that the average students' performance in problems involving the algebraic notation $\hat{\mathbf{i}}\,\hat{\mathbf{j}}\,\hat{\mathbf{k}}$ was better than the performance in the graphical (arrow) format, both in the context of mathematics and in the context of physics. Consistently, Liu and Kottegoda [86] highlighted a disconnect between undergraduates' understanding of the algebraic and geometric aspects of vectors.

## III. CONTEXT AND MOTIVATION FOR THE STUDY

As highlighted in the theoretical background, in research on students' difficulties in mathematics and in physics two main approaches can be distinguished [11,87]: *macroscopic* studies, aimed at identifying taxonomies of students' difficulties with relevant concepts or tools, and *microscopic* studies, grounded in theories of knowledge, aimed at describing students' knowledge and cognitive processes in much greater detail. Our study is of the first type. In order to understand the reasons for this choice and to interpret our results, we will now specify the context in which the research was designed.

The study presented here was conducted in 2018 at the University of Padua, a large-enrollment university in northern Italy. In the past few years, the Italian government has funded different actions aimed at sustaining students' enrollment in scientific degree programs and at preventing dropout [88]. In the context of such actions, in 2017 a survey was conducted among faculty who taught introductory physics courses. The aim was to gain insights into the main hurdles that students encounter in those courses, which are often a bottleneck in the students' career. The results of the survey suggested that one of the students' major difficulties was actually the use of mathematics. These results led to the development of a research project, the main goals of which were the design and administration of the instrument that we describe here, and the development of supporting actions based on its results. In particular, we identified the following research questions:

 (1) To what extent are students' difficulties in introductory physics due to difficulties with the mathematical tools that are considered prerequisite by instructors?

 (2) In what ways does students' performance in purely mathematical problems differ from their performance in parallel physics problems involving the same mathematical concepts?

Even if quantitative instruments exploring specific topics do exist in the literature [29,47,49,66,89,90], a "compact" instrument covering the most relevant mathematical topics for introductory physics courses while comparing the contexts of mathematics and physics was lacking. We think that our instrument actually fills this gap. We believe that this kind of assessment can be of interest from both a research and a practical point of view, since it allows testing students' difficulties across different mathematical topics using a single instrument, and it provides students and instructors with feedback that can be useful for improving their learning and teaching.

The project was supported by several departments belonging to the School of Science and the School of Engineering of our university and it involved 23 degree programs overall. As mentioned above, the expected impact was to better support both students and instructors in first-year physics courses.

## IV. METHODS

### A. Choice of the instrument

Since we wanted to survey a large number of students, we opted for a multiple-choice, distractor-driven test. Each of the multiple-choice questions contained in the test featured five options, only one of which was correct. Although this kind of instrument can bias students' responses by forcing their choice into one of the given options, it is useful when a large screening is sought.

It is worthwhile at this point to clarify our perspective on the relationship between mathematics and physics and on transfer. Though we acknowledge all the different perspectives on transfer, due to the choice of the instrument we do not expect our results to provide fine-grained information on perspectives such as PFL [30] or TiP [31]. In fact, at a first sight our design might look more similar to a traditional, "sequestered problem solving" setting [30]. However, in some studies, such as Ref. [35], research designs similar to ours have been interpreted according to modern perspectives such as AOT [32]. Though we believe that a full AOT account would require richer, qualitative data that would only be available through interviews, our study acknowledges this perspective in that its working hypothesis is that students may not necessarily see isomorphic problems as "similar" even if discipline experts would consider them as such. In other words, our research hypothesis is that students may adopt different approaches in solving "isomorphic" problems in a purely mathematical context and in a physical context. Our goal is to quantify these discrepancies and to give an account of the differences in students' answers in pairs of matched mathematics and physics problems in terms of the distractors that they choose more frequently.

### B. Test development

To design the test, we combined a literature review with an analysis of end-of-semester exams in order to select a number of relevant subtopics. Since representational fluency is relevant for expertise in physics, we decided to

TABLE I.　Summary of test items.

| Item | | Topic | | Representation | |
|---|---|---|---|---|---|
| Mathematics | Physics | Mathematics | Physics | Input | Output |
| 1M | 1P | Derivatives | Position → velocity | Words and formal | Words |
| 2M | 2P | Derivatives | Position → velocity | Graphs | Graphs |
| 3M | 3P | Derivatives | Position → velocity | Graphs | Numbers |
| 4M | 4P | Derivatives | Position → velocity | Words | Words and numbers |
| 5M | 5P | Derivatives | Position → velocity | Graphs | Words |
| 6M | 6P | Derivatives | Position → velocity | Words | Graphs |
| 7M | 7P | Integrals | Velocity → displacement | Graphs | Words |
| 8M | 8P | Integrals | Velocity → displacement | Graphs | Graphs |
| 9M | 9P | Integrals | Velocity → displacement | Graphs | Numbers |
| 10M | 10P | Vectors—components | Velocity (2D) | Words | Graphs |
| 11M | 11P | Vectors—components | Velocity (2D) | Graphs | Formal (unit vectors) |
| 12M | 12P | Vectors—components | Forces | Graphs | Formal (magnitude and angle) |
| 13M | 13P | Vectors—sum | Forces | Graphs | Formal (components) |
| 14M | 14P | Vectors—sum | Forces | Graphs | Formal (magnitude and angle) |
| 15M | 15P | Vectors—difference | Velocity (2D) | Graphs | Graphs |
| 16M | 16P | Vectors—dot product | Work | Graphs and formal | Numbers and formal |
| 17M | 17P | Vectors—cross product | Torque | Graphs and formal | Numbers and formal |

include different representations of the concepts covered in the test. We considered four broad classes of representations: words (labeled "W"), graphs (labeled "G"), formal language (algebraic expressions and equations, labeled "F") and numbers (labeled "N"). Each of these representations could be used either in the question ("input") or in the answers ("output"), or in both. For each combination of representations, two items were created: one item in the context of mathematics and one parallel item in the context of physics. Here we use the word context in its common meaning to distinguish purely mathematical items (labeled "M") from their physics counterpart (labeled "P"), without any reference to broader meanings. For the mathematical items, we used the formalism typically used in calculus courses, in order to highlight differences in framing the problems that might be due to notation issues. Some of the items that we designed were inspired by previous research, but we adapted and re-elaborated all the items to fit into the goals and structure of our assessment. Other items were specifically designed for the test. The distractors for each item were also built based on the literature, and/or on the analysis of our students' written exams.

An initial pool of 78 items were developed according to this logic and this preliminary ("pilot") version of the test was checked by experts (faculty who teach physics or calculus in introductory courses at our university) to assess content validity. The pilot test was administered in Spring 2017 to 71 first-year students enrolled in the degree course in architectural engineering, where one of the authors (G. T.) was lecturing, at the beginning of their physics course. An item analysis was performed, and students' responses were examined more in detail by conducting think-aloud interviews with 10 volunteer students from the sample. Based on the results, some of the items and/or the

distractors were deleted, added, modified, or rephrased. The final version of the test contains 34 items (17 in the context of mathematics +17 in the context of physics). The list and categorization of the items according to the context mathematics (labeled M), and physics (labeled P) and the representational forms used in each item are reported in Table I.

The test was administered to the students in Spring 2018, at the beginning of their physics course. Participation was voluntary but encouraged by the instructors. The test was delivered online using the *Moodle* platform of the Department of Physics and Astronomy of the University of Padua, setting a time limit of 90 min for completing the test. The sample consists of 1252 first-year students enrolled in 23 different degree programs belonging to the School of Science (35%) or to the School of Engineering (65%) of the University of Padua, Italy. 69% of the respondents were male. All of the students had followed a calculus course in the first semester [91]. At the time when the test was administered, 65% of the students had passed the calculus exam (of which 26% with high marks), 24% stated they had taken the exam but they had not passed it, and the remaining 11% had not yet taken the exam. Before entering university, 59% of the students had attended a "Liceo Scientifico" (scientific high school) and 38% had attended a technical school in the technological sector. Other types of schools were less represented (≤5%).

The full text of the TCV-MP is reported as Supplemental Material [92] both in English and Italian (the original language in which it was written and administered). The items are listed in the same order as they were delivered to the students (1M-2M-…-17M-1P-2P-…-17P, i.e., the whole mathematics part was given before the whole physics part).

For readability, in this paper we report the correct answer as option A for all the items, but the options were randomized in the version administered to the students.

## V. RESULTS

### A. Test mean score

The box plots describing the score distribution for the whole test, for the mathematics part, and for the physics part are shown in Fig. 1. The test mean score was 58%; the difference between the mean score in the mathematics part (61%) and the physics part (55%) was significant ($p < 0.001$) according to a paired two-tailed $t$ test, with effect size (Hedges $g$) $g_{av} = 0.25$ [93].

It is interesting to compare students' scores in the test with their score in the calculus exam. Their relationship is displayed in Fig. 2. We notice that the median of the test scores increases as the score in the calculus exam increases, but also that almost the entire range of test scores is covered for each of the score bands in the calculus exam.

### B. Instrument reliability and discriminatory power

We used some common statistical measures to assess the reliability and discriminatory power of our test [94]. In particular we evaluated the Kuder-Richardson index (KR20) as a measure of internal consistency of the test (reliability) and Ferguson's delta as a measure of the test's global discriminatory power. We obtained a KR20 of 0.91, indicating a good reliability of the test as a whole, and a Ferguson's delta of 0.99, suggesting that the test is also well discriminating. For the evaluation of individual items we employed three statistical measures: the facility index $FI$ (corresponding to the percent of correct answers normalized to 1), the point-biserial coefficient $r_{pb}$ (measuring item reliability defined as the correlation between the correctness of the item and the test score), and the discrimination index $DI_{27\%}$ (a measure of the item's ability to discriminate between the top-scoring students and the bottom-scoring
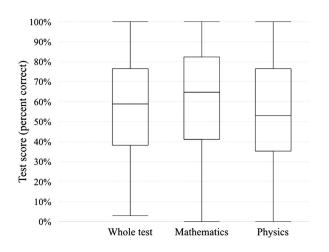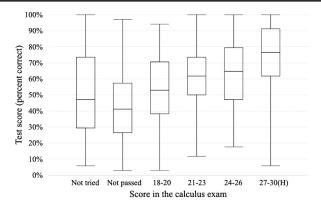


FIG. 2. The relationship between students' scores in the test and in the calculus exam they took in the first semester.

students; the percentage indicates that the groups were defined by the top to bottom 27%) [95].

Figure 3 displays the facility index of each item, comparing parallel items in the mathematics part and in the physics part of the test. As can be seen from the figure, facility indices range from 0.37 (item 16M) to 0.94 (item 7M) for the mathematics part, and from 0.32 (item 17P) to 0.83 (items 1P and 11P) for the physics part. The average facility index was 0.58, corresponding to the test mean score normalized to 1. The point-biserial coefficients range from 0.29 (item 11P) to 0.64 (item 2P), with an average of 0.50. All items fulfil the acceptability criterion $r_{pb} \geq 0.20$. The discrimination index was good ($\geq 0.40$) for fourteen items out of seventeen, with the maximum value for item 2P (0.81); it was acceptable for item 1P (0.35; a common acceptability criterion is $DI_{27\%} \geq 0.30$), while it was low for items 7M (0.19) and 11M (0.22). These low $DI_{27\%}$ values correspond to items with very high facility indices, i.e., almost all the students answered these items correctly.

The statistical indices relative to the individual items, together with the percentage of students who selected each of the five options for the different items, are reported in Table II for the context of mathematics and in Table III for the context of physics. A summary of the values of the different statistical indices used to evaluate test reliability is given in Table IV.

### C. Degree of association between isomorphic items

In Table V we compare pairs of matched items in mathematics and in physics by reporting (a) the difference in facility index, (b) phi coefficients ($\Phi$) for each pair of matched items. Phi coefficients, which were also used in similar studies [29] quantify the correlation between pairs of items; therefore, a low value of $\Phi$ ($0.10 \leq \Phi \leq 0.30$) indicates that students' performance in the two items was poorly correlated, whereas a high $\Phi$ value ($\geq 0.50$) indicates a highly correlated performance. High $\Phi$ values do not necessarily indicate that the students recognized the items as similar, but low $\Phi$ values suggest that the students saw the



FIG. 1. Box plots of students' scores for the whole test, for the mathematics part and for the physics part.
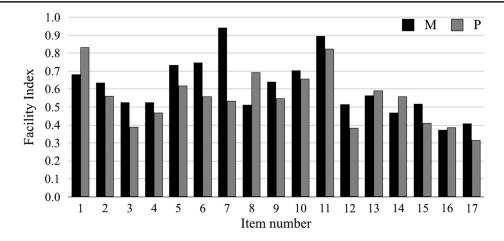
FIG. 3. Facility index vs item number in the two contexts of mathematics (labeled M) and of physics (labeled P).

items as different. In order to identify item pairs where the students' performance may be regarded as "different," we have labeled FI differences larger than 10% and $\Phi$ values smaller than 0.30 with an asterisk, and we have marked in bold the item pairs having at least one of the two values above threshold. A deeper understanding of these values requires a more detailed analysis of each item, which we report in the following.

### D. Item analysis

In the following, we compare students' answers to all pairs of matched items, both in terms of students'

TABLE II. Point-biserial coefficients ($r_{pb}$), discrimination indices ($DI_{27\%}$) and percentage of students who selected each of the five options for each item in the context of mathematics.[a]

| Item | $r_{pb}$ | $DI_{27\%}$ | A | B | C | D | E | Omit |
|---|---|---|---|---|---|---|---|---|
| 1M | 0.42 | 0.49 | 68 | 6 | 5 | 13 | 4 | 4 |
| 2M | 0.60 | 0.72 | 64 | 16 | 9 | 4 | 5 | 2 |
| 3M | 0.58 | 0.74 | 52 | 13 | 6 | 17 | 7 | 4 |
| 4M | 0.41 | 0.52 | 53 | 5 | 11 | 19 | 10 | 2 |
| 5M | 0.57 | 0.64 | 73 | 13 | 3 | 4 | 3 | 3 |
| 6M | 0.47 | 0.53 | 75 | 9 | 6 | 2 | 7 | 1 |
| 7M | 0.35 | 0.19 | 94 | 1 | 1 | 1 | 1 | 1 |
| 8M | 0.52 | 0.65 | 51 | 14 | 6 | 8 | 8[b] | 5 |
| 9M | 0.62 | 0.74 | 64 | 13 | 5 | 5 | 5 | 7 |
| 10M | 0.45 | 0.50 | 70 | 20 | 5 | 3 | 2 | 1 |
| 11M | 0.29 | 0.22 | 89 | 5 | 1 | 1 | 2 | 1 |
| 12M | 0.45 | 0.57 | 51 | 21 | 15 | 9 | 1 | 3 |
| 13M | 0.37 | 0.45 | 57 | 17 | 4 | 11 | 9 | 3 |
| 14M | 0.50 | 0.63 | 47 | 21 | 6 | 9 | 9 | 8 |
| 15M | 0.48 | 0.59 | 52 | 15 | 14 | 11 | 6 | 2 |
| 16M | 0.55 | 0.69 | 37 | 7 | 11 | 19 | 14 | 12 |
| 17M | 0.48 | 0.60 | 41 | 7 | 5 | 11 | 19 | 17 |

[a]Option A is the correct one for all the items; the order of the options was randomized in the version administered to the students. The value in the A column normalized to 1 gives the facility index.

[b]In this item there were six options. Option F was selected by the same percentage of students as option E (8%).

performance and in terms of the distractors that were more frequently chosen. In fact, as could already be inferred by looking at Tables II and III, relevant differences in the answer profile of parallel items were observed. We retrieved many of the typical mistakes that have been reported in the literature and in previous taxonomies, but we also report some new findings.

#### 1. Item 1

Item 1M contained an algebraic expression for the first derivative of a function, and the students had to relate the coefficients in this expression to the sign of the slope of the tangent line to the graph of the function. In the parallel

TABLE III. Point-biserial coefficients ($r_{pb}$), discrimination indices ($DI_{27\%}$) and percentage of students who selected each of the five options for each item in the context of physics.[a]

| Item | $r_{pb}$ | $DI_{27\%}$ | A | B | C | D | E | Omit |
|---|---|---|---|---|---|---|---|---|
| 1P | 0.39 | 0.35 | 83 | 5 | 2 | 2 | 3 | 5 |
| 2P | 0.64 | 0.81 | 56 | 17 | 10 | 9 | 4 | 3 |
| 3P | 0.63 | 0.78 | 39 | 5 | 6 | 45 | 2 | 3 |
| 4P | 0.44 | 0.56 | 47 | 8 | 3 | 12 | 26 | 4 |
| 5P | 0.59 | 0.74 | 62 | 22 | 2 | 1 | 10 | 3 |
| 6P | 0.59 | 0.75 | 56 | 17 | 12 | 2 | 10 | 3 |
| 7P | 0.56 | 0.73 | 53 | 14 | 5 | 8 | 13 | 7 |
| 8P | 0.56 | 0.64 | 69 | 5 | 3 | 13 | 6 | 4 |
| 9P | 0.55 | 0.70 | 55 | 17 | 4 | 15 | 3 | 5 |
| 10P | 0.55 | 0.66 | 66 | 16 | 4 | 4 | 6 | 5 |
| 11P | 0.48 | 0.43 | 83 | 3 | 3 | 2 | 2 | 7 |
| 12P | 0.48 | 0.57 | 39 | 20 | 15 | 13 | 2 | 11 |
| 13P | 0.50 | 0.59 | 59 | 10 | 4 | 8 | 9 | 10 |
| 14P | 0.54 | 0.68 | 56 | 9 | 6 | 4 | 10 | 14 |
| 15P | 0.51 | 0.62 | 41 | 22 | 24 | 2 | 2 | 9 |
| 16P | 0.53 | 0.65 | 39 | 7 | 11 | 12 | 12 | 19 |
| 17P | 0.42 | 0.50 | 32 | 5 | 10 | 13 | 14 | 26 |

[a]Option A is the correct one for all the items; the order of the options was randomized in the version administered to the students. The value in the A column normalized to 1 gives the facility index.

TABLE IV.   Summary of the values of the five statistical indices used to assess test reliability.

| Test statistic | Desired value | Value |
|---|---|---|
| Kuder-Richardson index | $>0.70$ | 0.91 |
| Ferguson's delta | $>0.90$ | 0.99 |
| Facility index (FI) | [0.30, 0.90] | av. 0.58 |
| Point-biserial coefficient ($r_{pb}$) | $\geq 0.20$ | av. 0.50 |
| Discrimination index (DI$_{27\%}$) | $\geq 0.30$ | av. 0.60 |

TABLE V.   Facility index (FI) differences and phi coefficients ($\Phi$) for pairs of parallel items in the context of mathematics (M) and in the context of physics (P). FI differences larger than 10% and $\Phi$ values smaller than 0.30 are labeled with an asterisk, and the item pairs having at least one of the two values above threshold are marked in bold.

| Item pair | FI(M) | FI(P) | FI difference (M-P) | $\Phi$ |
|---|---|---|---|---|
| **1** | 0.68 | 0.83 | −0.15* | 0.28* |
| 2 | 0.64 | 0.56 | +0.08 | 0.49 |
| **3** | 0.52 | 0.39 | +0.13* | 0.36 |
| **4** | 0.53 | 0.47 | +0.06 | 0.20* |
| **5** | 0.73 | 0.62 | +0.11* | 0.35 |
| **6** | 0.75 | 0.56 | +0.19* | 0.26* |
| **7** | 0.94 | 0.53 | +0.41* | 0.14* |
| **8** | 0.51 | 0.69 | −0.18* | 0.28* |
| 9 | 0.64 | 0.55 | +0.09 | 0.42 |
| 10 | 0.70 | 0.66 | +0.04 | 0.46 |
| **11** | 0.89 | 0.83 | +0.06 | 0.27* |
| **12** | 0.51 | 0.39 | +0.12* | 0.35 |
| 13 | 0.57 | 0.59 | −0.02 | 0.44 |
| 14 | 0.47 | 0.56 | −0.09 | 0.35 |
| **15** | 0.52 | 0.41 | +0.11* | 0.35 |
| 16 | 0.37 | 0.39 | −0.02 | 0.36 |
| 17 | 0.41 | 0.32 | +0.09 | 0.35 |

physics item (1P), an algebraic expression for the time derivative of position was given, and the students were asked to determine when the object's velocity was negative. The students performed much better on item 1P (83% correct) than on item 1M (68% correct), with $\Phi = 0.28$. For item 1M, the most common incorrect answer was distractor D, consisting in using the wrong coefficient to determine the sign of the slope.

### 2. Item 2

In item 2M, students were given the graph of a function and they had to select the correct graph representing the function's first derivative. In its parallel item 2P, the students were given a position-time graph and they had to identify the corresponding velocity-time graph. In both contexts, the input function was quadratic.

The students' performance was similar in the two items (64% correct in 2M, 56% correct in 2P, $\Phi = 0.49$). The most common incorrect answer in both contexts (16% for item

2M, 17% for item 2P) was distractor B, corresponding to a sort of "linearization" of the input graph. This kind of error was not included in previous taxonomies. One possible interpretation of this result follows Elby's "What You See Is What You Get" account [25]. According to this perspective, students are attracted by relevant perceptive features of a graph such as "going up-going down" and these features may take prevalence over less evident conceptual features. Similarly, Moore and Thompson [96] reported that students often adopt a "static shape thinking" in which graphs are interpreted as static objects that are described in terms of macroscopic trends rather than as representations of covarying quantities.

It could be argued that students following this kind of reasoning may also choose distractor D, corresponding to a graph which reproduces the input graph exactly. One of the possible reasons why the students prefer distractor B over distractor D could be that they probably know that the first derivative of a quadratic function is a linear function, which leads to the exclusion of distractor D. However, they cannot identify the exact relationship between the function and its first derivative, and as a way out, they opt for a graph that satisfies the condition of being linear, but at the same time reminds them of the original graph. This kind of reasoning was actually observed in some of the pilot interviews.

Finally, some students (10%) selected distractor C in item 2P, corresponding to a graph that is correct in the first two-thirds, but incorrect in the third one, where students' reasoning might have been similar to the WYSIWYG reasoning or to the "static thinking" reasoning described above. In fact, the transition from a function that "goes down" to a function that "goes up" in distractor C reflects the going up-going down behavior of the input graph in this region. Similar students' graphs were reported, for instance, by McDermott *et al.* [46] and by Bajracharya [78].

### 3. Item 3

In item 3M, the students were required to calculate the first derivative of a function at a point, given its graph. In the parallel physics item 3P, they had to calculate an object's velocity at a given instant, given its position-time graph.

The students performed better on item 3M (52% correct) than on item 3P (39% correct). The preferred incorrect answer for both contexts, distractor D, consists in calculating $y/x$ (or $s/t$) rather than $\Delta y/\Delta x$ (or $\Delta s/\Delta t$). This kind of mistake has been categorised in previous taxonomies as "non-origin slope error." However, the percentage of students who chose this distractor was dramatically different between the two contexts. In item 3M (context of mathematics), distractor D was selected by 17% of the students, while in item 3P (context of physics) 45% of the students chose it—even more students than the ones who selected the correct answer (39%). Another difference between the two contexts concerns distractor B, which corresponds to reading the value of the function directly off the $y$ axis. In item

3M, 13% of the students selected this answer, while in its parallel item 3P it was chosen by only 5% of the students. The mistake represented by this distractor was categorized as "confusion between slope and height" in previous taxonomies, although some authors such as Wemyss and Van Kampen [50] have argued that students who commit this mistake may not be really confusing the two variables, but rather just picking up the only answer they can think of.

These findings suggest that, in the context of physics, miscalculating the slope was only part of the problem. It is well known that students often define "velocity" as "space over time", an oversimplification that they have learned at school and that may be reinforced by the fact that, in high school, students mostly interact with position-time graphs that pass through the origin [50]. Our results confirm that issues related to the incorrect prior learning of physics cannot be underestimated. Consequently, instructors should be warned against considering students' mistakes in calculating velocities as a mere misapplication of mathematical knowledge and they should be aware that students may hold incorrect or oversimplified interpretations of physical concepts even when they have good results in mathematics.

### 4. Item 4

Items 4M and 4P concerned the relationship between the first derivative of a function and the function's maxima. The two parallel items were formulated a little differently. In item 4M (context of mathematics), the students were given information about the sign of a function's derivative and they had to decide where the function had its maximum value within the given interval. In item 4M (context of physics), information was given about the sign of an object's acceleration and the object's velocity at a point, and the students were required to choose the correct option describing the object's velocity at another point. A similar percentage of students answered these two items correctly in the two contexts (mathematics 53%, physics 47%), but the $\Phi$ coefficient was low (0.20), suggesting that the two performances are weakly correlated. In fact, by checking the students' answers more in detail, it turns out that the number of students who answered only one item (either M or P) correctly is comparable to the number of students who answered both items correctly or both items incorrectly.

### 5. Item 5

In item 5M, the students were required to relate the graph of a function to a verbal description of its first derivative. In its parallel item 5P, the students had to relate a position-time graph to a verbal description of the object's motion, formulated in terms of its velocity.

The students' performance was slightly better on item 5M (73% correct) than on item 5P (62% correct). The most common incorrect answer in both contexts was distractor B, but it was selected by a higher percentage of students in the context of physics (22%) than in the context of

mathematics (13%). According to previous taxonomies, this mistake could be classified as "variable confusion". Following this account, the students who selected this option did not pay attention to the fact that the given graph represented the function whereas the description was referred to the function's derivative, and they chose an answer that would be correct if both the graph and the description were both referred to the same variable. Another difference between the two items is that, in item 5P, a significant portion of the students (10%) chose distractor E, i.e., they interpreted the horizontal axis as the line along which the motion occurs. Previous research by Trowbridge and McDermott [42] has highlighted that, in physics, students sometimes employ kinematics concepts indiscriminately. Difficulties in interpreting the direction of motion and separating the shape of a graph from the path of the motion have also been reported in the literature [46,50] and we may also consider Elby's WYSIWYG perspective [25] as an alternative interpretation of this kind of mistakes.

Students' choice of distractor B could also be interpreted using the notion of "graphical forms," introduced by Rodriguez and colleagues [40,97]. Similarly to Sherin's symbolic forms [98], graphical forms involve associating intuitive mathematical ideas to a pattern, which in this case is a region of a graph. For example, the graphical form "steepness as rate" entails the idea that different levels of steepness in a graph correspond to different rates; "straight means constant" involves the idea that a straight line indicates a lack of change; and "curve means change" implies that a curve indicates a changing rate. Graphical forms are important in interpreting the "story" represented in a graph and are therefore particularly relevant for this pair of items, where students are required to associate a graph with its "story" told in verbal language. According to the graphical form account, in item 5M the graphical form "curve means change" and/or "straight means constant" may have been activated in response to the fact that line in the graph is not straight: this graphical feature may, in fact, have attracted the students towards distractor B, which contains the word "decreasing," rather than towards the correct answer, that contains the word "constant." Correspondingly, in item 5P the first part of the graph may have activated the graphical form "straight means constant," and the second part of the graph may have activated the graphical form "steepness as rate", associated with the words "constant" and "slows down" in distractor B.

### 6. Item 6

In item 6M, students were given a verbal description of a function in terms of the function itself and of its derivative, and they had to identify the correct graph representing the function. In its parallel item 6P, the motion of an object was described verbally and students had to select the corresponding position-time graph.

Item 6M was answered correctly by 75% of the students, while item 6P was answered correctly by 56% of the

students. The low $\Phi$ value (0.26) also suggests that the two items were seen as different. Similarly to the previous pair of items, the most common incorrect answer (distractor B) would be categorized as "variable confusion" according to previous taxonomies. However, this mistake can also be interpreted according to the graphical forms account. In fact, in item 6M the graphical form "straight means constant" may have been activated in association with the part of the graph where $x \geq 4$, and correspondingly, in item 6P, the word "constant" might have cued the students into choosing graphs that feature straight lines, like the ones in distractors B and E.

### 7. Item 7

Item 7M probed students' interpretation of the definite integral of a function as the area under the graph of the function, while its parallel item 7P probed the students' knowledge of the physical meaning of the area under a velocity-time graph.

Items 7M and 7P were the ones where the largest difference between the two contexts was observed: while in item 7M most of the students (94%) correctly associated the definite integral with the area under the curve, in item 7P only 53% of the students choose the correct answer. The small $\Phi$ value (0.14) confirms that the students actually saw the two items as different. The most common incorrect answers in item 7P was associating the area with the object's acceleration (14%), and saying that the area has no meaning in physics (13%). Both mistakes have been reported in the literature and they have been categorized as "area ignorance" in previous taxonomies [47].

### 8. Item 8

It item 8M, the students were given the graph of a function $f$; $F(x)$ was defined as the definite integral of the function from 0 to $x$, and the students had to identify a graph that could represent F. Correspondingly, in item 8P, students were given a velocity-time graph and they had to select the correct graph representing the object's displacement.

Against the trend, the students performed better in the context of physics (69%) than in the context of mathematics (51%), with a $\Phi$ value of 0.28. For sure, one of the reasons is that item 8M contained an extra distractor. However, the response profile was quite different in the two contexts, suggesting that the presence of the extra distractor does not fully explain the observed difference. In fact, in item 8M the most common incorrect choice was distractor B (a graph having the same shape as the input graph), while in item 8P the preferred distractor was D (13%), containing a graph that differs from the correct one for the sign of the curvature of the parabola in the first part. It may be that these students remembered that "if the velocity-time graph is linear, then the displacement-time graph is a parabola," without however being able to determine the sign of the curvature correctly. Another prominent reason for the observed difference is that

the figure in 8P is much simpler than the figure in 8M. In fact, the graph in 8M is continuously curving while the majority of the graph in 8P is not only straight, but horizontal. The graphs were chosen based on typical end-of-semester tests and the corresponding students' mistakes, but we recognize that a graph of the curve in 8M is more complicated to figure out than the integral of a constant.

### 9. Item 9

In item 9M, the students were given the graph of a function and they had to calculate the definite integral of the function up to a given point. In its parallel item 9P, students had to calculate an object's displacement up to a certain time instant, given its velocity-time graph.

Students' performance was similar in the two contexts (64% correct for item 9M, 55% for item 9P). The $\Phi$ value was 0.42. Items 9M and 9P were isomorphic except for a difference in the formalism. In fact, item 9M mentioned the definite integral of the given function up to point $x = 2$ as $F(2)$, whereas item 9P asked for the displacement "between $t = 0$ and $t = 2s$." This discrepancy might explain the observed difference in the answer profile. In fact, in item 9M the preferred distractor was B, corresponding to the value of the function at $x = 2$, while in item 9P 15% of the students choose distractor D, corresponding to the difference of the values of the function in $t = 0$ and in $t = 2$. The role of the different formalism used in mathematics and physics has been discussed in the literature [1].

### 10. Item 10

In items 10M and 10P, the students were given the verbal description of a vector and they to identify its graphical representation. In the purely mathematical version, the students had to identify a vector of magnitude 1 forming a positive angle with the $x$ axis, while in the parallel physics item the students had to identify the correct graphical representation of a velocity vector of magnitude 5 m/s along a direction forming a positive angle with the $x$ axis. A similar percent of correct answers was observed in the two contexts (70% for item 10M, 66% for item 10P, $\Phi = 0.46$). In both cases, the most common incorrect answer consisted in selecting a vector having both components equal to the given vector magnitude.

### 11. Item 11

In items 11M and 11P, the students were given a graphical representation of a vector and they had to identify its algebraic representation. In the context of physics, a velocity vector was given. The percent of correct answers was very high in both contexts (89% for item 11M, 83%). The low $\Phi$ value is due to the fact that the number of students who answered only item 11P correctly is similar and even higher than the number of students who answered both items incorrectly.

### 12. Item 12

In items 12M and 12P, the students had to select the correct algebraic expression for the components of a vector, starting from its graphical representation. In the context of physics, the given vector represented the weight of an object moving on an incline.

Consistent with the literature [70,71], students had more difficulties in the context of physics (39% correct) than in the purely mathematical context (51% correct). In both contexts, the most common mistake consisted in choosing the wrong sign for the negative component of the given vector (20%–21%), and in inverting the sine and the cosine of the given angle (15%). This latter mistake was categorized as "confusion between sine and cosine" in previous taxonomies, but it could also be due to the fact that the angle indicated as $\theta$ was the complement of the angle used in the formulas for components, which is also commonly labeled $\theta$. The choice of this angle was intentional, since we wanted to detect situations where students use a remembered formula rather than looking at the graph for context. Remarkably, in item 12P a relevant percentage of the students (11%) did not provide an answer.

### 13. Item 13

In item 13M, the students were given two input vectors ($\vec{A}$ and $\vec{B}$) in graphical form, and they had to select the correct option for the magnitude of their sum, expressed algebraically. In the corresponding physics problem, two forces acting on an object were displayed graphically, and the students were required to calculate the magnitude of the net force on the object. The percentage of correct answers in the two contexts was basically the same (57% for item 13M, 59% for item 13P) and the $\Phi$ value was 0.44. However, in item 13P, the students who chose an incorrect answer were almost equally spread across the different options (including skipping the question), while in item 13M the majority of the students who selected an incorrect answer (17%) chose option B, corresponding to the wrong sign of the $x$ component of one of the input vectors.

### 14. Item 14

In item 14M, two input vectors ($\vec{A}$ and $\vec{B}$) were displayed, and the students had to select the correct option for the $x$-component of their sum, expressed algebraically. In the context of physics, two forces acting on an object were displayed, and students had to identify the correct algebraic representation of the $x$ component of the net force. Students' performance in the two items was similar (47% correct on item 14M, 56% correct on item 14P). The $\Phi$ value was 0.35. The slightly worse performance in mathematics could be due to the fact that, in item 14M, one of the two input vectors ($\vec{B}$) had a negative $x$ component; choosing the incorrect sign for the $x$ component of $\vec{B}$ was

indeed the most common incorrect choice (21%) in this item.

### 15. Item 15

Items 15M and 15P were about students' vector difference, and the vectors were represented graphically both in the question and in the answers. The physical context for item 15P was the motion of the Moon around Earth: two vectors representing the velocity of the Moon at two instants of time were displayed, and the students had to select the vector corresponding to the change in velocity over the given time interval.

Item 15M was answered correctly by 52% of the students, while item 15P was answered correctly by only 41% of the students. The most common mistakes were calculating the sum of the two vectors instead of their difference (distractor B, selected by 15% of the students in item 15M and by 22% of the students in item 15P), and choosing a vector of magnitude equal to the magnitude difference of the two input vectors (distractor C, selected by 14% of the students in item 15M and by 24% of the students in item 15P, where the two input vectors had the same magnitude).

In order to highlight the role that contextualization might have played in the students' framing of these two items, we report an excerpt from one of the pilot interviews. Just before this conversation, the student had tried to solve item 15M, and he had calculated the sum of the two vectors instead of their difference. Then he was asked to solve item 15P.

> *Student: It is about circular motion. We have centripetal acceleration. Centripetal acceleration is… v square over r… it points to the center… [looks puzzled]*
> *Interviewer: What are you thinking about?*
> *S: None of these answers is correct, because it [acceleration] points to the center, so it should be like this [draws a vector pointing to the center, starting at point 1] here, and like this [draws a vector pointing to the center, starting at point 2] here. The Moon moves on a circle, but its speed is constant. This is constant circular motion.*
> *I: Ok. If I told you the correct answer was there, what would you say?*
> *S: Well, if I must choose among these ones, I'd say zero [distractor C], since they [the two vectors] have the same magnitude. But there is centripetal acceleration, unless they are asking something different.*

The student did not immediately frame item 15P as a problem about vector difference. Instead, he started recalling miscellaneous facts and formulas about circular motion, which, however, did not cue him towards the correct answer. When invited to select one of the given options, the student chose option C (corresponding to the magnitude

difference) since it was the only thing he could think of, although he was not convinced. Though the students' knowledge of the physics of the problem could have guided him towards the correct answer, in fact it mainly distracted him by activating unhelpful resources. Rather than a "failure" in transfer, this situation could be seen as a different framing of the problem depending on the context: since the student expected the problem to be about circular motion, he activated his resources accordingly. According to a transfer in pieces perspective [31], the student has consistently applied his knowledge resources in precisely those situations where they were relevant according to his framing.

### 16. Items 16 and 17

Items 16M/16P concerned the dot product, while items 17M/17P concerned the cross product. The physical context for the dot product was the work done by a force along a given path, while the cross product was contextualized as the torque exerted by a force on a bar.

The percent of correct answers for these two pairs of items were among the lowest ones in the test (37% for item 16M and 41% for item 16P; 39% for item 17M and 32% for item 17P). In items 16M/16P, the most common mistake was exchanging the sine and the cosine, while the second most common mistake was interpreting the result of the dot product as a vector in the same plane as the two input vectors. In items 17M/17P, the most common incorrect answers consisted in selecting a vector in the same plane as the input ones and in calculating the dot product instead of the cross product. Just like in items 12M/P, we cannot infer from the results if the confusion between sine and cosine was "genuine" or if it was rather due—at least in part—to the choice of the wrong angle. In any case, students committing this kind of mistake were not using geometric reasoning: they were engaging in a purely "calculation mode" rather than trying to make sense of the problem.

The low percent of correct answers in these two pairs of items can be interpreted as a confirmation that vector products are a very difficult topic for students. It is remarkable that a large percentage of students actually skipped these items, particularly in the context of physics. This fact could be seen as a further indication of students' difficulty with this topic. However, we should also take into account that the students had just started their study of physics at the university level. In fact, though work and torque are covered in high school, they are often not treated as vector products before university, and it might be that the students who skipped these questions were simply not familiar with this description.

### E. Comparison across different representations

As mentioned above, we designed the test items including different representations of the concepts covered by the test, in order to see if the different representations played a role in determining students' problem solving ability. We have actually found evidence of such differences, as it can be seen by comparing the facility indices of items on the same topic (e.g., derivatives) containing different representations. For example, facility indices range from 0.52 (item 3M) to 0.75 (item 6M) across the six items about derivatives in the context of mathematics, and from 0.39 (item 3P) to 0.83 (item 1P) across the corresponding items in the context of physics. This suggests that students' ability in answering problems about derivatives also depends on the representational formats used in the problem.

The comparison between items 2P and 3P provides an excellent example of how students' performance can be inconsistent across the different representations. The two items were about the relationship between position and velocity. Item 2P was solved correctly by 56% of the students, while only 39% of them selected the correct answer in item 3P. The input for both items was an object's position-time graph, but the output representational format was different: in item 2P, the students were asked to select the object's velocity-time graph, while in item 3P they had to calculate the object's velocity at a given instant of time numerically.

Below we analyze the two items from the point of view of a student by reporting the following excerpt from her think-aloud interview.

*Student: I think the answer is A.*
*Interviewer: Ok. Can you explain your reasoning?*
*S: I take the tangent for each point [she mimics the tangent with her hand]. At the beginning the slope is positive. Then it goes like this [puts her hand horizontally], so the derivative is zero. And then it goes like this [mimics a negatively sloped line]. Negative. Therefore, the answer is A.*
*I: Ok. What about item 3P?*
*S: [She reads it aloud]. "The velocity…." Er… I don't know. I mean it is… the derivative… but the derivative of what? I am used to see x…x to the square…*
*I: You referred to the derivative correctly when solving item 2P. What did you use in that case?*
*S: The tangent line.*
*I: Can you use this idea to solve item 3P?*
*S: No. I mean, I can figure out how the straight line behaves, but… how do I calculate the numbers? I need an equation to calculate the numbers.*
*I: Since you do not have any equations, can you think of anything else you could do?*
*S: I don't know. Maybe I should make a general calculation and so it would be… space over time… [she hesitates]*
*I: You're not convinced, are you?*
*S: No, because that holds when it [the velocity] is constant, but it is not constant here.*

Apparently, this student understands the geometrical meaning of the first derivative of a function, and she is able to use this concept to go from the graph of a function to the graph of its derivative. She can also use this idea correctly in the context of physics, recognizing velocity as the time derivative of position and mentioning both the graphical (slope of the tangent) and functional (first derivative) relationship between the two kinematical quantities. However, when asked to calculate a point derivative numerically, the student fails to do that, claiming that she needs an equation to calculate numbers. It seems that the student is looking for a function to match, i.e., she is engaging in a purely "calculational" game, or, equivalently, seeking for a *plug-and-chug* solution rather than trying to make sense of the problem [56,78]. She is thinking of position and velocity algebraically, without any reference to the physical meaning of these two quantities. According to Rodriguez *et al.* [40], this student would be described as a "nonblender": she never refers to physics ideas or to "blended" ideas, but only to mathematical concepts and procedures. It is interesting to notice that even in item 2P (which she solved correctly) the student adopted a nonblended reasoning, which was, however, productive in that case.

## VI. DISCUSSION AND CONCLUSIONS

In this paper we have described an assessment, named "test of calculus and vectors in mathematics and physics," which we developed with the aim of comparing students' ability to answer questions on derivatives, integrals, and vectors in a purely mathematical context and in the context of physics. The assessment, a multiple-choice test, included multiple representations (graphs, words, numbers, and formal expressions) of the chosen concepts. We now discuss how the results presented above provide insights into our research questions.

Our first research question was *To what extent are students' difficulties in introductory physics due to difficulties with the mathematical tools that are considered prerequisite by instructors?*

Although there is an overall correlation between students' performance in the mathematics part and in the physics part of the test, the effect is not particularly strong. Also the comparison between students' score in the test and their score in the calculus exam suggests that it is not sufficient to rely only on calculus courses for learning how to use mathematics in the context of physics. Focusing our inquiry down to individual pairs of parallel items, the analysis of $\Phi$ coefficients suggests that students do not necessarily use the same strategy on parallel questions in the context of mathematics and in the context of physics, or in other words, they sometimes frame "parallel" isomorphic differently depending on the context. The degree of association between parallel questions depends on the specific subtopic and on the representations used in the

problem. In some cases, students seem to have context-specific preferred strategies which are likely to depend on previous instruction.

The qualitative analysis of students' answers leads us to our second research question: *In what ways does students' performance in purely mathematical problems differ from their performance in parallel physics problems involving the same mathematical concepts?*

For a relevant number of items, the students' answer profile (frequency of choice of each option) in pairs of matched items in the two contexts was quite different, even when the percent of correct answers was similar. This means that the students' attention was directed towards different distractors in a context-specific manner. In context-rich problems, students are easily distracted by irrelevant features that can activate unhelpful resources, but even when visual cues suggest a similarity between parallel questions in mathematics and in physics, students sometimes apply different strategies. Particularly for some topics, the acquisition of solid context-specific procedures (e.g., using formulas to calculate quantities in kinematics) seems to be difficult to discard and it may limit the students' problem solving abilities. It is therefore confirmed that previous instruction plays a very relevant role in enhancing or hindering students' capacity to use different strategies. Often, students tend to be stuck in a purely calculational mode, rather than trying to make sense of the problems. Finally, the choice of a specific strategy also depends on the representations and/or the formalism used in the question and in the answers.

In general, our results confirm that a good performance in physics is not just a matter of knowing enough math. Often, the students' main difficulty does not lie in the mathematics itself, but rather in constructing a "blended" mathematics-physics framework [2]. Sometimes, students may develop "synthetic" approaches that can be considered a first step towards the blending of physics and mathematics, but this process is not automatic and not always productive. As a consequence for instruction, we argue in favor of explicitly training the students in the mathematization of physics as a competence, highlighting the structural role of mathematics, rather than relying on single, automated problem-solving procedures where mathematics only plays a technical role. This suggestion is consistent with authors such as Uhden *et al.* [3], whose mathematical-physical model for the use of mathematics in physics education supports the adoption of teaching strategies focussed on the structural dimension.

The relationship between mathematics and physics is particularly relevant for university instruction, where the two disciplines are taught separately by faculty belonging to different departments and having different backgrounds. Bridging the gap between the two contexts is often a responsibility that is left to the students. Although we may agree that students of this age should have a reasonable

degree of autonomy and responsibility for their learning, we think that a deeper collaboration between mathematics and physics instructors should be sought. In particular, we recommend to our colleagues who teach physics in introductory courses not to take the mathematization of physics for granted, and we encourage them to provide explicit instruction on it. A possible, favorable context for this specific instruction could be recitations held by teaching assistants (TAs). TAs have the opportunity to meet students in smaller groups, and they could help the students reflect on the ways they frame and solve problems and support them in developing a blended reasoning.

We think that the test we have described here can be a useful instrument for both instructors and students. We believe that the added value of this instrument is that it considers different aspects of previous research, addressing three topics that are very relevant for physics. Using the test, instructors and TAs can obtain a "picture" of their classroom in order to adjust their teaching and to address their students' difficulties more effectively, whereas the students can get individual feedback on their preparation. Finally, we think that our results can be relevant for other researchers in physics education who are interested in similar topics.

A future development of the project will be to design online learning modules that will be offered to the students based on their results in the test. The content of the modules will follow the topics of the test, while the activities will be focused on mathematization and on the use of different representations in physics. Moreover, we are testing a modified version of the test to be used in secondary schools with the aim of improving students' mathematization skills before they enter university.

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

[2] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend conceptual and formal mathematical reasoning in solving physics problems, Sci. Educ. **97**, 32 (2013).

[3] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, Sci. Educ. **21**, 485 (2012).

[4] G. Pospiech, Framework of mathematization in physics from a teaching perspective, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer International Publishing, London, 2019), pp. 1–33.

[5] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. **5**, 020108 (2009).

[6] R. Karam, O. Uhden, and D. Httecke, The "math as prerequisite" illusion: Historical considerations and implications for physics teaching, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer International Publishing, London, 2019) pp. 37–52.

[7] J. R. Hundhausen and R. Yeatts, An experiment in integration: Calculus and physics for freshmen, J. Eng. Educ. **84**, 369 (1995).

[8] J. W. Dunn and J. Barbanel, One model for an integrated math/physics course focusing on electricity and magnetism and related calculus topics, Am. J. Phys. **68**, 749 (2000).

[9] C. Michelsen, Expanding the domain: Variables and functions in an interdisciplinary context between mathematics and physics, in *Proceedings of the 1st International Symposium of Mathematics and its Connections to the Arts and Sciences*, edited by A. Beckmann, C. Michelsen, and B. Sriraman (Verlag Franzbecker, Hildensheim, 2005), pp. 201–214.

[10] E. F. Redish, Problem solving and the use of math in physics courses, in *Proceedings of "World View on Physics Education in 2005: Focusing on Change"* (2006), http://www.physics.umd.edu/perg/papers/redish/IndiaMath.pdf.

[11] E. F. Redish, A theoretical framework for physics education research: Modeling student thinking, in *Physics Education Research, Proceedings of the Varenna Summer School, "Enrico Fermi" Course CLVI, Italian Physical Society*, edited by E. F. Redish and M. Vicentini (Società Italiana di Fisica, Bologna, 2004) pp. 1–64.

[12] J. Tuminaro, A cognitive framework for analyzing and describing introductory students' use and understanding of mathematics in physics, Ph.D. Thesis, University of Maryland, 2004.

[13] L. Cui, N. S. Rebello, P. R. Fletcher, and A. G. Bennett, Transfer of learning from college calculus to physics courses, in *Proceedings of the NARST 2006 Annual Meeting* (2006), https://web.phys.ksu.edu/papers/2006/Cui_NARST2006.pdf.

[14] F. R. Yeatts and J. R. Hundhausen, Calculus and physics: Challenges at the interface, Am. J. Phys. **60,** 716 (1992).

[15] H. T. Hudson and W. R. McIntire, Correlation between mathematical skills and success in physics, Am. J. Phys. **45,** 470 (1977).

[16] J. Tuminaro and E. F. Redish, Elements of a cognitive model of physics problem solving: Epistemic games, Phys. Rev. ST Phys. Educ. Res. **3,** 020101 (2007).

[17] D. Domert, J. Airey, C. Linder, and R. Lippmann Kung, An exploration of university physics students' epistemological mindsets towards the understanding of physics equations, NorDiNa **3,** 15 (2007).

[18] S. M. Barnett and S. J. Ceci, When and where do we apply what we learn? A taxonomy for far transfer, Psychol. Bull. **128,** 612 (2002).

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

[20] S. Britton, P. B. New, A. L. Roberts, M. D. Sharma, A. Brew, and J. Sachs, Investigating students' ability to transfer mathematics, in *Transforming a University: The Scholarship of Teaching and Learning in Practice* (Sydney University Press, Sydney, Australia, 2007).

[21] N. S. Rebello, L. Cui, A. G. Bennett, D. Zollman, and D. J. Ozimek, Transfer of learning in problem solving in the context of mathematics and physics, in *Learning to Solve Complex Scientific Problems*, edited by D. H. Jonassen (Routledge, New York, 2017), Chap. 10.

[22] J. Lobato, Alternative perspectives on the transfer of learning: History, issues, and challenges for future research, J. Learn. Sci. **15,** 431 (2006).

[23] J. Lobato and B. Rhodehamel, "Noticing" as an alternative transfer of learning process, J. Learn. Sci. **21,** 433 (2012).

[24] J. Lobato, C. Hohensee, and B. Rhodehamel, Students' mathematical noticing, J. Res. Math. Educ. **44,** 809 (2013).

[25] A. Elby, What students' learning of representations tells us about constructivism, J. Math. Behav. **19,** 481 (2000).

[26] J. F. Wagner, A Transfer-in-Pieces consideration of the perception of structure in the transfer of learning, J. Learn. Sci. **19,** 443 (2010).

[27] A. M. Madsen, A. M. Larson, L. C. Loschky, and N. S. Rebello, Differences in visual attention between those who correctly and incorrectly answer physics problems, Phys. Rev. ST Phys. Educ. Res. **8,** 010122 (2012).

[28] J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, The cost of concreteness: The effect of nonessential information on analogical transfer, J. Exper. Psychol. **19,** 14 (2013).

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

[30] J. D. Bransford and D. L. Schwartz, Rethinking transfer: A simple proposal with multiple implications, in *Review of Research in Education*, edited by A. Iran-Nejad and P. D. Pearson (American Educational Research Association, Washington, DC, 1999), Vol. 24, pp. 61–100.

[31] J. F. Wagner, Transfer in Pieces, Cognit. Instr. **24,** 1 (2006).

[32] J. Lobato, The Actor-Oriented Transfer perspective and its contributions to educational research and practice, Educ. Psychol. **47,** 232 (2012).

[33] T. J. Nokes, Mechanisms of knowledge transfer, Think. Reas. **15,** 1 (2009).

[34] N. S. Rebello, Consolidating traditional and contemporary perspectives of transfer of learning: A framework and implications, in *Proceedings of the NARST 2007 Annual Meeting* (2009), https://web.phys.ksu.edu/papers/2007/rebello-narst.pdf.

[35] L. Cui, Assessing college students' retention and transfer from calculus to physics, Ph.D. Thesis, Kansas State University, 2006.

[36] N. S. Rebello and D. Zollman, A model for dynamic transfer of learning, in *Proceedings of the NARST 2005 Annual Meeting* (2005), https://www.researchgate.net/publication/255580937_A_Model_for_Dynamic_Transfer_of_Learning.

[37] D. Hu and N. S. Rebello, Using conceptual blending to describe how students use mathematical integrals in physics, Phys. Rev. ST Phys. Educ. Res. **9,** 020118 (2013).

[38] T. J. Bing and E. F. Redish, The cognitive blending of mathematics and physics knowledge, AIP Conf. Proc. **883,** 26 (2007).

[39] K. Bain, J.-M. G. Rodriguez, A. Moon, and M. H. Towns, The characterization of cognitive processes involved in chemical kinetics using a blended processing framework, Chem. Educ. Res. Pract. **19,** 617 (2018).

[40] J.-M. G. Rodriguez, S. Santos-Diaz, K. Bain, and M. H. Towns, Using symbolic and graphical forms to analyze students' mathematical reasoning in chemical kinetics, J. Chem. Educ. **95,** 2114 (2018).

[41] P. W. Thompson, The development of the concept of speed and its relationship to concepts of rate, in *The Development of Multiplicative Reasoning in the Learning of Mathematics*, edited by G. Harel and J. Confrey (SUNY Press, Albany, NY, 1994), pp. 181–234.

[42] 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).

[43] 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).

[44] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics Education Research, Am. J. Phys. **67,** 755 (1999).

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

[46] 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).

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

[48] G. Zavala, S. Tejeda, P. Barniol, and R. J. Beichner, Modifying the test of understanding graphs in kinematics, Phys. Rev. Phys. Educ. Res. **13,** 020111 (2017).

[49] A. Dominguez, P. Barniol, and G. Zavala, Test of understanding graphs in calculus: Test of students' interpretation of calculus graphs, Eurasia J. Math. Sci. Technol. Educ. **13,** 6507 (2017).

[50] 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).

[51] 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).

[52] D.-H. Nguyen and N. S. Rebello, Students' understanding and application of the area under the curve concept in physics problems, Phys. Rev. ST Phys. Educ. Res. **7,** 010112 (2011).

[53] P. W. Thompson and J. Silverman, The concept of accumulation in calculus, in *Making the connection: Research and Teaching in Undergraduate Mathematics Education* (Mathematical Association of America, Oberlin, OH, 2008), pp. 43–52.

[54] 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).

[55] S. R. Jones, Understanding the integral: Students' symbolic forms, J. Math. Behav. **32,** 122 (2013).

[56] S. R. Jones, Areas, anti-derivatives, and adding up pieces: Definite integrals in pure mathematics and applied science contexts, J. Math. Behav. **38,** 9 (2015).

[57] S. R. Jones, The prevalence of area-under-a-curve and anti-derivative conceptions over Riemann sum-based conceptions in students' explanations of definite integrals, Int. J. Math. Educ. Sci. Technol. **46,** 721 (2015).

[58] S. R. Jones, Prototype images in mathematics education: The case of the graphical representation of the definite integral, Educ. Studies Math. **97,** 215 (2018).

[59] R. R. Bajracharya, T. M. Wemyss, and J. R. Thompson, Student interpretation of the signs of definite integrals using graphical representations, AIP Conf. Proc. **1413,** 111 (2012).

[60] 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).

[61] 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).

[62] A. Susac, A. Bubic, E. Kazotti, M. Planinic, and M. Palmovic, Student understanding of graph slope and area under a graph: A comparison of physics and non-physics students, Phys. Rev. Phys. Educ. Res. **14,** 020109 (2018).

[63] M. Planinic, A. Susac, L. Ivanjek, and Ž. Milin Šipuš, Comparing student understanding of graphs in physics and mathematics, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer International Publishing, London, 2019), pp. 233–246.

[64] R. Knight, The vector knowledge of beginning physics students, Phys. Teach. **33,** 74 (1995).

[65] N.-L. Nguyen and D. E. Meltzer, Initial understanding of vector concepts among students in introductory physics courses, Am. J. Phys. **71,** 630 (2003).

[66] P. Barniol and G. Zavala, Test of understanding of vectors: A reliable multiple-choice vector concept test, Phys. Rev. ST Phys. Educ. Res. **10,** 010121 (2014).

[67] A. Susac, M. Planinic, D. Klemencic, and Z. Milin Sipus, Using the Rasch model to analyze the test of understanding of vectors, Phys. Rev. Phys. Educ. Res. **14,** 023101 (2018).

[68] S. Flores, S. E. Kanim, and C. H. Kautz, Student use of vectors in introductory mechanics, Am. J. Phys. **72,** 460 (2004).

[69] P. Shaffer and L. C. McDermott, A research-based approach to improving student understanding of the vector nature of kinematical concepts, Am. J. Phys. **73,** 921 (2005).

[70] J. Van Deventer and M. C. Wittmann, Comparing student use of mathematical and physical vector representations, AIP Conf. Proc. **951,** 208 (2007).

[71] J. Van Deventer, Comparing student performance on isomorphic math and physics vector representations, Master's Thesis, The University of Maine, 2008.

[72] A. Van Heuvelen, Learning to think like a physicist: A review of research based instructional strategies, Am. J. Phys. **59,** 891 (1991).

[73] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2,** 020103 (2006).

[74] P. Klein, A. Müller, and J. Kuhn, Assessment of representational competence in kinematics, Phys. Rev. Phys. Educ. Res. **13,** 010132 (2017).

[75] 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. **22,** 22 (2014).

[76] P. Nieminen, A. Savinainen, and J. Viiri, Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning, Phys. Rev. ST Phys. Educ. Res. **8,** 010123 (2012).

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

[78] R. R. Bajracharya, Student application of the fundamental theorem of calculus with graphical representations in mathematics and physics, Ph.D. Thesis, The University of Maine, 2014.

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

[80] 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).

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

[82] P. Nieminen, A. Savinainen, and J. Viiri, Force Concept Inventory-based multiple-choice test for investigating students' representational consistency, Phys. Rev. ST Phys. Educ. Res. **6,** 020109 (2010).

[83] M.-A. Geyer and W. Kuske-Janen, Mathematical representations in physics lessons, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer International Publishing, London, 2019), pp. 75–102.

[84] S. Van den Eynde, P. van Kampen, W. Van Dooren, and M. De Cock, Translating between graphs and equations: The influence of context, direction of translation, and function type, Phys. Rev. Phys. Educ. Res. **15,** 020113 (2019).

[85] 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).

[86] D. Liu and Y. Kottegoda, Disconnect between undergraduates' understanding of the algebraic and geometric aspects of vectors, European J. Phys. **40,** 035702 (2019).

[87] M. D. Caballero, B. R. Wilcox, L. Doughty, and S. J. Pollock, Unpacking students' use of mathematics in upper-division physics: Where do we go from here?, Eur. J. Phys. **36,** 065004 (2015).

[88] M. Carli, S. Hemmer, S. Lippiello, and O. Pantano, *Il Tutorato Formativo: Un modello di tutoring integrato per l'empowerment degli studenti universitari* (PensaMultimedia, Lecce, 2018), Chap. X.

[89] J. Epstein, The Calculus Concept Inventory: Measurement of the effect of teaching methodology in mathematics, Not. Am. Math. Soc. **60,** 1018 (2013).

[90] J. Epstein, Development and validation of the calculus concept inventory, in *Proceedings of the 9th International Conference on Mathematics Education in a Global Community* (2007), http://math.unipa.it/~grim/21_project/21_charlotte_EpsteinPaperEdit.pdf.

[91] Students at our university take a final exam on calculus at the end of the first semester and they start physics in the second semester of their first year.

[92] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.16.010111 for the full text of the TCV-MP both in English and Italian.

[93] D. Lakens, Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas, Front. Psychol. **4,** 863 (2013).

[94] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5,** 020103 (2009).

[95] T. L. Kelley, The selection of upper and lower groups for the validation of test items, J. Educ. Psychol. **30,** 17 (1939).

[96] K. C. Moore and P. W. Thompson, Shape thinking and students' graphing activity, in *Proceedings of the 18th Meeting of the MAA Special Interest Group on Research in Undergraduate Mathematics Education*, edited by T. Fukawa-Connelly, N. E. Infante, K. Keene, and M. Zandieh (RUME, Pittsburgh, PA, 2015), pp. 782–789.

[97] J.-M. G. Rodriguez, K. Bain, M. H. Towns, M. Elmgren, and F. M. Ho, Covariational reasoning and mathematical narratives: Investigating students' understanding of graphs in chemical kinetics, Chem. Educ. Res. Pract. **20,** 107 (2019).

[98] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19,** 479 (2001).