Using smartphones and tablets as experimental tools in the physics classroom: Effects on learning and motivation

Alice Gasparini,[1,2] F. Stern,[1] M. Delaval,[3] and A. Müller[1,2]

[1]*Institute of Teacher Education, University of Geneva, Geneva, Switzerland*
[2]*Department of Physics, University of Geneva, Geneva, Switzerland*
[3]*Université de Lille, ULR 4072—PSITEC—Psychologie: Interactions, Temps, Emotions, Cognition, Lille, France*

(Received 20 August 2024; accepted 10 December 2025; published 11 March 2026)

Tablets and smartphones provide compact and relatively inexpensive tools for measuring many physical quantities, and numerous applications of these "mobile devices as experimental tools" (MDETs) have been published over the last decade, covering a broad range of topics. According to the literature, MDETs can offer educational benefits by creating authentic, real-life contexts for physics learning, enhancing student motivation through the use of familiar technology, and supporting cognitive processes by providing multiple representations (videos, tables, and graphs) of physical phenomena. On the other hand, serious concerns have been raised about potential distractions and cognitive overload. Regarding these conflicting perspectives, few empirical studies on the impact of MDETs in real classroom settings of regular, full-length physics courses are available, and—to the best of our knowledge—none focus on a nonspecialized high-school target group. The purpose of this contribution is to present a study of a mechanics course in such a setting, addressing the tight curricular, material, and practical constraints inherent to it. A quasiexperimental pre- and post-test design comparing a treatment group using MDETs and a control group without (same content, lesson plan, and teacher) was used. The 19-week teaching sequence focused on conceptual learning and motivational outcomes (e.g., interest, self-concept, and curiosity), controlled by several predictor variables. Findings reveal substantial pre- and post-test learning gains for both groups (Cohen's $d = 0.9$) and small gains for perceived relation to reality ($d = 0.29$). However, no significant differences between treatments were found, indicating that the use of MDETs does not outperform conventional approaches under the given constraints. On the other hand, no evidence of negative effects such as distraction or cognitive overload was observed. Moreover, little to no interactions with predictors such as gender or prior knowledge were found. In conclusion, MDETs show considerable potential as a viable and effective option for integrating technology into physics teaching, offering learning outcomes comparable to those of successful conventional teaching, but not better. Furthermore, no evidence of negative effects from distraction or cognitive overload was found, nor of strong interactions with gender and other predictors. The MDETs approach thus appears suitable for both sexes and different kinds of learners. This leads to suggestions for future utilization of MDETs as a component of physics education, which are discussed as potential perspectives.

DOI: 10.1103/96gp-gqvj

I. INTRODUCTION

Tablets and smartphones are easy to handle and widespread among people of all generations; their use as learning tools in science classes is nowadays a reality [1,2]. The built-in sensors provide a new kind of compact and relatively inexpensive instruments for many physical quantities. In physics education, a broad range of applications of these mobile devices as experimental tools (MDETs[1]) have been published since about a decade, covering mechanics [3,4], acoustics and waves [5–10], optics [11,12], magnetism [13,14], radioactivity [15,16], and other fields. For reviews see Refs. [1,17]; for a recent broad collection see Ref. [2]. Phyphox (Physical Phone Experiments) is a highly versatile, user-friendly open-source software for sensor control, readout, and data export

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.

[1]To the best of our knowledge, the acronym M(D)ET was coined by Kuhn and Vogt [3], together with similar ones (e.g., NET = new experimental tool).

for use with MDETs [18]. A first advantage of using MDETs in physics education is the possibility to perform relatively (with respect to traditional classroom instruments) effortless and at the same time precise measurements, supporting a real-time interpretation of the data, both in the classroom or in the everyday-life situations. This makes it possible for learners to learn science in a "topical" context (a situation) that is perceived as more authentic than traditional sciences courses, e.g., a vertical jump as an application of Newtonian mechanics and energy conservation [4]; see also the numerous examples cited above. Moreover, as these devices are part of everyday life of learners, their utilization during physics lessons may provide an authentic "material" context (an instrument) for them [19]. This authenticity matches the framework of context-based science education [20,21] based on theoretical and empirical arguments that providing meaningful, real-life contexts can have a positive impact on motivation and learning (see Sec. II A). However, a positive effect on motivation implies only an indirect effect for learning, and other existing research implies conflicting arguments for direct effects of MDETs regarding physics learning (see Secs. II B and II C). Moreover, the empirical evidence of the direct effects of smartphone use on physics learning seems to be inconclusive at present [22]. On one hand, they represent powerful instruments whose apps can easily and straightforwardly produce tables, diagrams, and many other functions that—in a traditional course—have to be produced by the learners themselves. This can lead to improved learning, as already shown for other information and communications technology (ICT) applications to physics education [23]. Yet other studies also indicated negative effects of ICT on learning, in particular, distraction and cognitive overload, impeding learning [24–26].

Specifically regarding MDETs, many experiments have been proposed in the last decade (see references above), but only a few empirical studies about their educational impact exist, which were either carried out in single lab sessions (about 3 h; [19]), among advanced learners [27–30], or both [31]. The goal of our investigation is thus to examine the effects of the use of MDETs in physics education, taking into account and completing the previous results. The main differences with respect to previous studies are the systematic integration into a whole learning sequence in regular physics classroom, at high school level, and the duration of the intervention (36 lessons of 45 min).

II. RESEARCH BACKGROUND

Prior studies have established a well-developed research background for the use of MDETs in physics education, integrating elements of context-based science education, cognitive-affective theory of multimedia learning, and other elements from research in science education and educational science. Two further well-established areas of physics education research on which we draw are conceptual learning and the connection between mathematics and physics learning. We review each of these strands of research with a focus on elements relevant for the present study.

A. Context-based science learning

As stated above, a remarkable advantage of using MDETs in science classes is the possibility to perform new types of activities, creating new situations that would not be possible using a traditional lab setup, whether inside or outside the classroom. Thus, new contexts for science learning can be conceived, closer to students' everyday lives and their real interests. This matches the theoretical framework of context-based science education (CBSE; [20,21]), positing that science learning related to actual, real(istic), genuine contexts and experiences learners are supposed to encounter can foster their motivational and cognitive activation and thus lead to improved attitudes to and understanding of science. Indeed, such positive effects on attitudes and motivation (with large effect sizes) and learning (with medium effect sizes) have been revealed in some empirical studies [20,32,33]. The approach of CBSE is also strongly advocated by Program for International Student Assessment (PISA), see, e.g., the statement of the importance of learning activities "that could be part of the actual experience or practice of the participant in some real-world setting [and] places most value on tasks that could be encountered in a variety of real-world situations" [34] or by the in-depth analysis of Fensham [21]: "[R]eal world contexts have […] been a central feature of the OECD's PISA project for the assessment of scientific literacy among young people." This focus aligns well with the importance of real-life phenomena and real-life problems within the "Next Generation Science Standards" in the United States [35].

Despite encouraging empirical evidence and recommendations in favor of context-based science learning, a main obstacle to its widespread integration into science teaching practice stems from the difficulties encountered in preparing such courses in the long term, requiring a substantial adaptation of the existing pedagogical methods with respect to the actual curricula [36]. In other words, it is often difficult for teachers to reconcile the educational potential of CBSE with practical and curricular requirements, particularly over the long term.

MDETs constitute a new educational tool allowing teachers to reconcile the practice of context-based activities with existing educational constraints and settings. As argued by Hochberg *et al.* [19], they can provide a learning context in two different and related ways: First, as mentioned above, the flexibility and the "hands-on" format of the MDET activities opens up the possibility of new contexts for science learning, perceived as relevant to the everyday life of learners, whether inside or outside the school. For example, activities whose context is sport or a hobby practiced by the student: this is called a

topical context. Topical contexts are known to be able to foster interest and learning in science [20]. Second, smartphones and tablets are widespread and often have an important emotional value to young people: therefore, the use of these "hands-on" and attractive devices as measuring instruments during the physics lessons is an element of authenticity in itself, constituting what is called a *material context* (the device) [19]. Moreover, this exploits the attractiveness of new technologies to strengthen the interest of students in the studied topics, as shown by Swarat *et al.* [37]. However, an interest based only on the material context linked to mobile devices needs to be maintained in order to be transferred to the subject matter on a long-term basis, which is unlikely in the context of one-off activities, but rather requires a prolonged use of MDETs.

Among the two types of context (topical, material) possible with MDETs, the topical context is likely to promote a positive attitude toward sciences in a more direct way, because it is based on a more explicit link between the studied discipline and the center of interest of learners. Nevertheless, the effect of the topical context is reinforced if the material context is also present. Carrying out activities that combine both contexts, material and topical, is a promising approach to make the most of the potential of MDETs in sciences lessons [19]. A main purpose of the present work is to implement and study MDETs in the real classroom setting of a regular high school mechanics course.

Focusing on motivational outcomes and following the above reasoning, two important subdimensions of motivation might directly be impacted by the use of MDETs in the physics lessons and will be object of our investigation: the learners' *perception of the relation to reality* as well as the *interest in relation to the learning subject*. *Perceived relation to reality* reflects the extent to which the subjects treated in the course appear relevant and linked to the daily life of the learners and therefore it corresponds also to a measure of the perceived authenticity. *Interest* is defined as "a psychological state of engaging […] that occurs and develops during interactions between persons and their environment as objects, topics, contents, or ideas; interest is characterized by focused attention and a motivational reaction, such as enjoyment" [38–40]. Another key internal state of motivation in our investigation is the construct of *curiosity as a state*—that is, the curiosity toward the subject of learning—defined as "a desire for new information or experience that includes a trigger, a reaction to that trigger, and a resolution, which can be satisfactory or unsatisfactory." Indeed, if the curiosity of a learner is triggered, for example, by a stimulating context provided by an activity in the physics course, this can be the source of her or his wish to seek new information and to engage in a further learning, renewing in turn the initial state of curiosity [41].

In addition, the relative ease of access to the data and their representations makes it possible to minimize the effort of the learner, provided that she or he masters the functionalities of the apps used during the activities. Thus, the perceived competence of a learner to quickly, easily, and successfully conduct MDET activities can be increased, and this, in turn, can benefit both motivation and learning. This aspect is commonly understood in terms of *(situational) self-concept*, i.e., "a mental representation of oneself, including a collection of cognitive beliefs about oneself formed through experience and feedback from the environment" [42,43]. We thus include *situational self-concept* experienced during a learning sequence as a dependent variable in our study.

B. Cognitive (-affective) theory of multimedia learning

In Sec. II A, we explained how the creation of a meaningful context can foster motivation in science learning. Despite the widespread belief that a positive effect on motivation can lead to a positive effect on learning, meta-analyses have only shown a moderate correlation (0.3–0.4) between motivation and learning [44,45]. As a consequence, based solely on the effects that MDETs can have on motivation, one can only expect moderate *indirect* effects on learning. However, further arguments and empirical observations exist, regarding possible *direct* effects of MDETs on learning. These are framed in the *cognitive-affective*[2] *theory of learning with media* (CATML; [47,48]). This theory expands on the *cognitive theory of multimedia learning* (CTML; [49–51]) by integrating affective factors in the learning process, especially regarding attention allocation and cognitive engagement. Common to both theories is a strong foundation in cognitive load theory [52] and theories of multiple representations (dual coding theory [53] and extensions to further modalities [47]).

C(A)TML is based on several assumptions derived from prior research [46,48,54]: (i) humans process different information modalities through different channels; (ii) processing of current information occurs in short term (or working) memory with limited capacity for each channel; (iii) storage takes place in long-term memory, an active, evolving system, which holds and reproduces information on the basis of various mental processes (structuring, inference mechanisms, emotional tags); short- and long-term memory interact in both directions; (iv) meaningful learning takes place when learners are actively engaged in cognitive processes such as selecting, organizing (both within and between channels), and integrating new and existing information (retrieved from long-term memory); and (v) affective factors can mediate aspects such as attention allocation in the short term memory or active engagement in cognitive processes [related to

[2]"Affective" is sometimes used as an umbrella term for emotional and motivational dimensions, including interest and self-beliefs [46].

assumptions (ii) and (iv), respectively].[3] Assumptions (i) through (iv) are common to both CTML and CATML, while assumption (v) is specific to the latter.

One can see how multiple representations are related to assumptions (i), (ii), and (iv); physics learning, in particular, systematically requires simultaneous understanding, e.g., of a real experimental setting, a schematic representation, diagrams, tables, and mathematical expressions [55–57]. Cognitive load theory, in turn, comes into play in assumptions (ii) and (iv). Specifically, two aspects are of particular interest here: First, the concept of "element interactivity" as key factor for cognitive load (CL), where an "element" is a concept or a procedure to be learned and "interactivity" represents the degree of linkage between the elements necessary for a given learning task. Learning distinct elements with only a few connections (such as simply memorizing the three Newton's laws) is considered a "low element interactivity" task. In contrast, a task has a "strong element interactivity" when there are many connections between the different elements, for example, solving complex problems or carrying out physics experiments, where the links between the observations and the theory have to be understood and investigated. With the many multiple representations occurring naturally when learning and practicing physics (see above), element interactivity and thus CL are high. Second, the differentiation between "intrinsic" and "extraneous" cognitive load: *intrinsic cognitive load* is generated by the nature of the learning content and is inherent to the complexity of the learning subject itself (in particular due to high element interactivity, [58]), while *extraneous cognitive load* arises from the presentation of the content to be taught, in particular surface features, and does not contribute to learning. Henceforth, extraneous load should be reduced as much as possible to keep the limited short-term memory capacity available to deal with the intrinsic cognitive load. CATML emphasizes that design and presentation features of media that can lower cognitive load, thus helping to enhance learning outcomes. We will now apply this framework to MDETs as an educational approach in the physics classroom.

C. MDETs as an educational approach in the physics classroom

MDETs represent powerful instruments with the potential to reduce the extraneous CL during experiments, as the apps can easily and straightforwardly provide multiple representations and many other functions that—in a traditional course—have to be produced by the learners themselves. An example of multiple representations of the motion of the standing vertical jump ("squat jump") [4] provided by video analysis apps is shown in Fig. 1. These representations can be used, for example, to calculate the maximum jump height using various foundational relations of Newtonian mechanics. Other examples of similar kind can be found in collections [1,2]. These examples demonstrate how a considerable amount and diversity of data on motion experiments can be easily stored and analyzed within the mobile device, depending on learning objectives and level of learner. These possibilities leave more time and attention to interpret, classify, and organize relevant information and work out the element interactivity.

Based on its theoretical background and empirical evidence supporting it, C(A)TML has developed specific guidelines and design principles for instructional media [46,54]. The following principles are of special importance with respect to MDETs [31]:

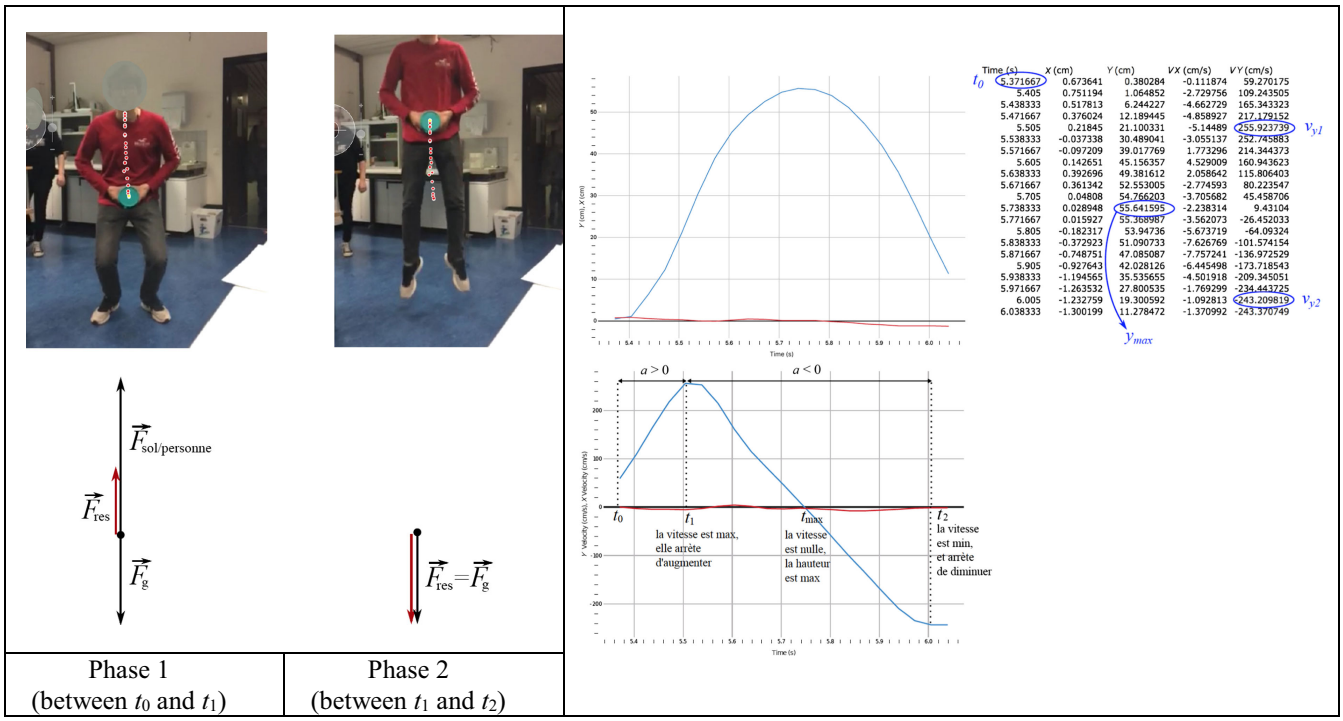
The principle of contiguity (spatial and temporal) for multiple representations states that (i) spatial proximity helps minimizing visual search processes, which can lead to a reduction in extraneous CL; (ii) temporal proximity helps avoiding additional CL that arises when mental representations need to be held active for longer durations. The above example (Fig. 1) shows how the simultaneous display of corresponding and complementary representations possible with MDETs fulfills the requirement of spatial and (by "swiping" rapidly between apps) also of temporal continuity.

The principle of interactivity (or segmenting principle) states that better learning occurs when learners can explore dynamic visualizations with sequence and pace determined by themselves. With MDETs, learners have full control over the different representations (real motion, stroboscopic images, tables, and diagrams) and can, for instance, switch back and forth between a stroboscopic image and the velocity diagram in order to understand their connection (an essential example of element interactivity for the learning of mechanics).

Finally, the positively moderating effect of motivation on attention allocation and active engagement in the learning process with MDETs was already discussed above.

Yet there are also plausible arguments that MDETs might actually not be conducive to learning. First, one might say that the offload of cognitive load in producing multiple representations (e.g., motion diagrams) is in fact counterproductive, since it is essential for learners to engage in this process themselves. Indeed, learning with multiple representations is known to be a critical issue in science learning, often creating considerable difficulties [59–61]. A solid understanding lies above all in the ability to make connections between the different representations and to create coherent connections within each representational format, that is to say to create intra- and inter-representational connections [51,62,63]. According to Mayer [51,63,64], these connections and the consistency between them are essential for effective learning, and the positive impact of

[3]These mediation factors are probably related to the indirect effects of motivation on learning discussed above. However (to the best of our knowledge), available research has not analyzed this question in detail.



Time (s)	X (cm)	Y (cm)	VX (cm/s)	VY (cm/s)
5.371667	0.673641	0.380284	-0.111874	59.270175
5.405	0.751194	1.064852	-2.729756	109.243505
5.438333	0.517813	6.244227	-4.662729	165.343323
5.471667	0.376024	12.189445	-4.858927	217.179152
5.505	0.21845	21.100331	-5.14489	255.923739
5.538333	-0.037338	30.489041	-3.055137	252.745883
5.571667	-0.097209	39.017769	1.773296	214.344373
5.605	0.142651	45.156357	4.529009	160.943623
5.638333	0.392696	49.381612	2.058642	115.806403
5.671667	0.361342	52.553005	-2.774593	80.223547
5.705	0.04808	54.766203	-3.705682	45.458706
5.738333	0.028948	55.641595	-2.238314	9.43104
5.771667	0.015927	55.368987	-3.562073	-26.452033
5.805	-0.182317	53.94736	-5.673719	-64.09324
5.838333	-0.372923	51.090733	-7.626769	-101.574154
5.871667	-0.748751	47.085087	-7.757241	-136.972529
5.905	-0.927643	42.028126	-6.445498	-173.718543
5.938333	-1.194565	35.535655	-4.501918	-209.345051
5.971667	-1.263532	27.800535	-1.769299	-234.443725
6.005	-1.232759	19.300592	-1.092813	-243.209819
6.038333	-1.300199	11.278472	-1.370992	-243.370749

FIG. 1. Multiple representations of the motion of a standing vertical jump, provided by smartphone video analysis apps: phase 1, stand-up (acceleration) process; phase 2, free motion under gravity [4]. Left: stroboscopic image of the trajectory and force diagram; right: graphs of position and velocity (y direction) as functions of time, table with x and y components of position and velocity. This is one of the labwork activities of the MDETs intervention (see Sec. III C).

actively processing different representations and their connections has been shown empirically [57]. This is the reason why caution is necessary when introducing these multiple representations into science learning sequences: learners must have the necessary prerequisites for handling all the representations provided [e.g., in mechanics diagrams for $x(t)$, $v(t)$]. Hochberg *et al.* [31] rightly state that a "blanket statement on how digital media can be meaningfully used in the classroom is not possible," and whether the facilitated use of multiple representations via MDETs is actually such a meaningful use is an open question.

Second, it is important to note that the above benefits are only possible for learners who, beyond sufficient understanding of relevant representations (as just stated), can additionally and rapidly handle the necessary functionalities using the device. Otherwise their use may turn out to be counterproductive and overload working memory, thus preventing learners from focusing on the content to be learned. Moreover, distracting effects of ICT in general can be observed, if its use in the classroom is not sufficiently structured (see, e.g., Ref. [65] for a study regarding laptops). Learners can make inappropriate use of the devices, for instance, by processing information that is not correctly connected to the learning purpose or by just "playing around," which is likely to lead to increased extraneous CL, as well as reduced attention and cognitive engagement. On the other hand, several more recent studies did not find such distracting effects of MDETs when used for physics learning, and they rather reported positive effects [19,27,31]. In fact, gaining empirical evidence on this interplay of positive and negative effects is a main research objective of this research.

D. Purpose and research questions

Even though many experiments using MDETs have been proposed in the last decade (see Sec. I), only few empirical studies about their educational impact exist (see Table I). Hochberg *et al.* [19,31] used MDETs with upper secondary students in a single learning unit on mechanics (pendulum). The results of their first study showed higher interest levels in learners of the treatment group (with a small to medium effect size), as well as a higher curiosity related to the studied subjects (small effect size; no differences in learning were observed [19]). In contrast, the second study among learners of a specialized physics course reports significantly higher levels of learning in the treatment group (with medium effect size), while no differences have been observed for motivational outcomes and cognitive load [31]. None of these two studies used a topical context, which leaves unanswered questions regarding the effects of using MDETs when exploiting its full potential, i.e., combining the topical and the material context. In another investigation in an undergraduate mechanics course, lasting a full semester and with both material and the topical

TABLE I. Summary of the main features of recent studies on the effects of MDETs in mechanics courses. *Within the constraints of a regular high school mechanics course matched to the official study plan.

Study	Educational level	Specialized target group	Duration of the intervention	Topical context	Material context
[19]	Secondary II	No	Single lab session	No	Yes
[31]	Secondary II	Yes	Single lab session	No	Yes
[27,66,67]	Undergraduate	Yes	One semester	Yes	Yes
This study	Secondary II	No	One semester	Yes*	Yes

contextualization taken into account, Klein *et al.* [27,28] obtained medium to large positive effects for conceptual understanding, relation to reality, curiosity, disciplinary authenticity, self-concept, interest, and autonomy, consistent with other work in the same target group [66,67]. Note that the studies having found positive effects on learning were among learner selections with higher levels of motivation and understanding (specialized high school [31] or undergraduate students [27]).

Today, to the best of our knowledge, there is no study on the use of MDETs in the real classroom setting of a regular mechanics course for upper secondary level students not specializing in physics and covering a whole teaching sequence matched to the official study plan. Note that such a course faces tight constraints, limiting feasible experiments and topical contexts. A main purpose of the present work is to present such an implementation of MDETs, combining material and topical context as far as possible within the given limitations.

The choice of learners not specializing in physics implies a focus on conceptual learning as educational objective, addressing learning difficulties and misconceptions well known from prior research in mechanics learning [68–71] (Chap. 14.1.2), including established assessment instruments (Ref. [72], Sec. III).

In this perspective, we produced a series of MDET activities, which will suitably replace the traditional laboratory activities in the treatment group during a whole semester, while maintaining the same educational objectives, teaching program, and organizational as well as temporal framework as the control group (Sec. III C). In addition, a test for conceptual learning in mechanics, based on the prior research just mentioned, and aligned with the local study program was developed and revalidated on the basis of existing, well-validated tests (Sec. III D 1). Furthermore, for assessing motivational outcomes, we utilized well-established instruments from the literature (Sec. III D 2).

On the basis of the above research background and our own development for the intervention and instruments, our research questions are as follows: What are

1. the cognitive effects (conceptual learning, cognitive load) and
2. the motivational effects (interest, perceived relation to reality, self-concept, and curiosity related to the studied topics)

of a MDET-based teaching sequence on mechanics for high school learners, as compared to a sequence without MDETs?

III. METHODS

A. Sample and setting

The study was carried out on 16- to 17-year-old learners from Geneva high school, having physics class groups with a maximum size of 16 learners. A sample breakup is given in Table II; 102 learners participated in the pilot study (PS) belonging to seven physics class groups taught by three different teachers, and 111 learners participated in the main study (MS), belonging to eight class groups (four pairs) taught by four different teachers. In the main study, each teacher taught one control and one treatment class.

The pilot study was run to validate instruments and the feasibility of the intervention. In particular, a conceptual learning questionnaire, adapted to the specific learning content of the intervention and the local curriculum, was newly developed (see Sec. III D 1), and several questionnaires for predictor and motivational variables were translated to French (see Secs. III D 2 and III D 3). An item analysis was carried out on these questionnaires (see

TABLE II. Sample characteristics of the study. Each participating teacher taught at least one control and one treatment class group. Sample sizes are given as total number (N_{tot})/number of female (N_f)/number of male (N_m) participants; CG, TG: control and treatment group.

	TG	CG	TG + CG	Teachers	
	$N_{tot}/N_f/N_m$	$N_{tot}/N_f/N_m$	$N_{tot}/N_f/N_m$	$N_{tot}/N_f/N_m$	Class groups
Pilot study (PS)	59/39/20	43/23/20	102/62/40	3/2/1	7
Main study (MS)	55/33/22	56/31/25	111/64/47	4/2/2	8

TABLE III. Timeline of the sequence and tests. Standard tests refer to the assessment tests made by the teachers, independently of the study; they contribute to the final average grade of physics.

Week	Part	Treatment group	Control group
1		Pretests	
2–5	1	Common lesson structure: introduction on physics background, paper-pencil exercises, corrections and answers for exercises and questions	
		MDET lab sessions and exercises	Conventional lab sessions and exercises
6		Standard test	
7–10	2	Schedule identical to chapter 1	
11		Standard test	
12	3	Schedule identical to chapter 1	
13 and 14		Exam session	
15–18	4	Schedule identical to chapter 1	
19		Post-tests	

below). Given the fairly good progress and achievement of the pilot study, the main study could be conducted in the same way, with minor adaptations to the teaching program and in the MDET activities. Pre- and post-questionnaires also worked well during the first study and could also be used up to minor adaptations.

Two further details have to be mentioned: first, as the main study took place during the school year 2019–2020, group 3 (see Table II) had to take the post-tests remotely during the first week of confinement due to the COVID pandemic. Second, the learners of group 4 were in their third year over a total of four years of upper secondary school, while the learners of the other groups were in their second year. While the program of the mechanics course was exactly the same, these students had more mathematics and physics (on other topics) than the rest of the sample.

B. Study and course design

The design was a repeated-measures control-treatment group comparison (control group vs treatment group: without vs with MDETs), which took place during learners' regular physics lessons. The intervention lasted a whole semester of 19 weeks. The teaching sequence followed the official teaching program,[4] with the following content:

(1) Introduction to kinematics: position vector, displacement, trajectory, velocity vector (average and instantaneous), speed (average and instantaneous), acceleration vector (mean and instantaneous).
(2) One dimensional kinematics: uniform linear motion (ULM), uniformly accelerated linear motion (UALM), free fall.
(3) Dynamics: Newton's laws 1, 2, 3, and the law of universal gravitation.

[4]https://www.ge.ch/document/programmes-disciplines-enseignees-dans-filiere-gymnasiale-au-college-geneve-valable-eleves-scolarises-2021.

The course design is shown in Table III. There were 13 weeks of classes, each consisting of two physics lessons, plus six weeks of testing and exams. In the treatment group, six activities based on MDETs replaced traditional experiments and exercises (i.e., there was the same number of experiments in both groups); the lab sessions had equal duration for the control group and the treatment group. Up to the lab activities, the lessons had a common structure (introduction for physics background, paper-pencil exercises, corrections and answers for exercises and questions) identical for the control group and the treatment group, and the learning content was also identical. Both groups were taught by the same four teachers. The precise order of these activities varied according to the teaching practices of each teacher, but the total proportion of time dedicated to each type of activity was the same for the control group and the treatment group across all class groups. Moreover, the same topical context was provided in both the control group and treatment group.

When, for practical reasons, the conditions did not allow the control group to carry out an experiment with the same topical context as treatment group (see example below), a conventional experiment was carried and the topical context in question was presented in an exercise sheet. All the participating teachers successfully completed the sequence and the test sessions, both in the PS and in the MS.

C. Learning materials

The MDET activities were designed and carried out by the researchers and the participating teachers, according to the educational objectives, the level of the learners, and the constraints related to the curriculum and the school setting. Based on the experiences of the PS, a few adaptations to the chronology and protocols of the activities were made in the MS. Particular care was taken to align the educational objectives with the contents and time constraints set by the school programs, as well as with conceptual obstacles for

the learning domain known from research (see Sec. III D 1). In particular, although the apps directly provide velocities, motion graphs, and regression analysis, the use of these features was introduced gradually within the sequence of activities (see Appendix B). Students only made use of them once they had performed the calculations and produced the graphs themselves, ensuring they had understood the meaning of all the physical quantities and the relationship between the observed motion, the associated functions, and their mathematical and graphical representation.

The activities were carried out using eight iPads provided to each treatment class group, in order to have at least one tablet for a pair of learners during the activity sessions. The applications "Video Physics" and "Graphical Analysis" by Vernier were used[5].

An example of a MDETs labwork is shown in Fig. 1, the standing vertical jump [4]. The process is divided into two main phases: The first one is the stand-up (acceleration) phase, when the feet are still in contact with the ground (between t_0 and t_1 in Fig. 1). Here the acceleration and the resultant net force are both constant and upward. The constant acceleration, both in the first and in the second phase, is found by calculating the slope of the velocity-to-time graph. The learners could draw the force diagram and had to apply their knowledge of dynamics (gravitation and the Newton's 2nd and 3rd laws) in order to find the pushing force of the person on the ground. The second phase (between t_1 and t_2 in Fig. 1) is the free motion under gravity (first upward, then downward) with a change in the orientation of the velocity, starting when the feet are no longer in contact with the ground. This kind of motion is central in the mechanics curriculum, as it allows the learner to overcome the difficulty of differentiating velocity (which changes orientation during movement) and acceleration (which is constant and opposed to initial velocity). Moreover, the link between the acceleration and the resulting force is also established in the experiment.

More examples of MDET activities are provided in the Appendix B, and a detailed description of all the activities, protocols, and worksheets is available in Ref. [73]. The following comment about the selection of MDET activities is in order. We emphasize that a main purpose of our work was to implement and study MDETs in the real classroom setting of a regular high school mechanics course, with all its curricular, material, and practical constraints. This implies limitations on the kinds of the student experiments that can be carried out and the topical contexts that can be realized (e.g., no amusement park physics, no project labs—at least not in the local curriculum). Nevertheless, the material context was combined with a topical context wherever conditions allowed it (in particular the "round trip" and "jump" activity, see below); moreover, the aspect of "ownership of data" is present in all activities.

[5]https://www.vernier.com/product/video-physics-for-ios/.

D. Variables and instruments

In this section, we will present the choice of variables and the construction of instruments. A more detailed description of the instruments' construction process, as well as all the questionnaires and tests administered, is available in Ref. [73].

1. Conceptual learning

Conceptual learning was assessed pre- and post-test using a multiple-choice questionnaire adapted from the Force Concept Inventory (FCI [74]), the Test of Understanding Graphs in Kinematics (TUG [75]), and the Motion Concept Test, (MCT [76]). Analogous to the sequence's learning activities, the concept learning test was designed within the curriculum, aligned with the research questions, and targeting common misconceptions. Each item was taken from or based on instruments already existing and validated in the literature. When no items addressed specific concepts, new questions aligned with the sequence's learning content were created and validated before the PS. The test of the PS was subsequently adapted and refined for the MS. The majority of the items were about one-dimensional kinematics, which represents most of the program covered during the duration of the study. Among others, questions about the interpretation of a stroboscopic image of a movement were included. These interpretations were particularly relevant in this study, because the tracing of a motion by the MDETs' apps provides exactly this kind of real-time representations, and therefore a difference in the results between the control group and the treatment group could be expected. Then, items testing misconceptions on vector displacement, average velocity, and average acceleration in a two-dimensional motion, on free fall, on average speed in a round trip, on the difference between velocity and acceleration, and on the three Newton's laws were included. The concept instrument of the MS consisted of 16 items in the pretest and 19 items in the post-test, scored with 1 point for a unique right answer. The total score was the total ratio of right answers. The result of the conceptual learning post-test (LPO) was included in the calculation of the learners' final physics grade, whereas the result of the conceptual learning pretest (LPR) was not. In order to motivate learners to apply themselves to sincerely answer the pretest questions, they were explained that the results of these tests would be important to assess their progress in the course in view of a better implementation of teaching.

2. Motivational variables

We studied the impact of the use of MDETs on four motivational variables: interest, perceived relation to reality, self-concept, and curiosity state. Instruments (six-level

Likert scales) were based on well-validated instruments [77–81], further developed and validated in Refs. [19,32]. After translation to French and revalidation in the pilot study, each final subscale contained between 5 and 7 items (on a 6-level Likert scale; 1 = completely disagree, 6 = completely agree). Sample items and results of the revalidation are reported in Sec. IV B.

3. Predictor variables

A set of potentially influential predictors with well-validated instruments is available from prior research in the field [19,27,31]. We summarize key features for predictors for cognitive and affective outcomes: rationale, sample items, and original sources, where more details can be found (abbreviations for later use in tables are also included). In addition to the prevalues of each outcome variable, six predictors related to learner characteristics were taken into account:

- *Prior physics grade* (PY_GRADE_PRE) and *prior mathematics grade* (MA_GRADE_PRE), which serve as indicators of initial general knowledge in the two disciplines.
- *Prior grade in the teaching language* (French), since all the activities, exercises, and the content of the physics course are based on the understanding of spoken and written language (FR_GRADE_PRE).
- *Spatial abilities* (SA), assessed by the "paper folding" test well validated from previous research [82] (p. 176), and discussed as important factor for science learning with multiple representations [83], in particular, in abstract domains like physics [84].
- *Curiosity as a trait* (CT), based on [80,81]; serves as baseline for curiosity state (Sec. III D 2), one of the motivational outcomes. Sample item: "I find it fascinating to learn new things."
- *Self-concept regarding the use of smartphone apps* (SCS; based on and adapted from [19,85–87]), because individuals with a positive self-concept might show higher motivation and performance in learning with smartphones. Sample item: "I am good using the apps for smartphones."

Five more predictors related to characteristics of learning and classroom processes were considered:

- *Cognitive load related to the experiments* and *to the use of the app for smartphones* (CLE and CLS, based on [88–90] and validated by [91–93]), as high CL might be harmful for both learning and motivation (Sec. II B). Sample items: "I could well focus on the experiments, without struggling with the equipment" (CLE) and "I could well focus on the experiments, without struggling with the apps for the smartphone" (CLS).
- *Cognitive activation during the experiments* (CAE; based on [94]).
- *Involvement during the physics lessons* (INV), as perceived by students; based on and adapted from [95,96] (WIHIC). Both CAE and INV belong to main positive mechanisms hypothesized for context-based science education (see Sec. II A). Sample items: "I was actively involved in doing the experiments" (CAE) and "I expressed my thoughts during class discussions" (INV).
- Assessment (by students) of the *teacher engagement* (TA), also based and adapted from [95,96] (WIHIC) in order to control for potential teacher difference between the treatment groups. Sample item: "The teacher's explanations helped me to understand."

All but two of the above instruments were used and validated in prior research on MDETs use in physics labs [19,31]. Within the available testing time (45 min in total) the two additional tests which could be included are those for *cognitive activation during the experiments and involvement during the physics lesson*, also well validated by prior research (see above). All tests had had satisfactory to good psychometric properties (e.g., all internal consistencies $\alpha_C \geq 0.75$, see Refs. [19,94,95]). After translation to French and revalidation in the pilot study, final subscales contained 5 to 7 items (on a 6-level Likert scale; 1 = completely disagree, 6 = completely agree). Results of the revalidation are reported in Sec. IV B.

Additionally, analyses were carried out for gender and for different class groups (i.e., a certain class taught by a certain teacher). For analysis of this set of predictors, see Sec. III E 2.

E. Data analysis

1. Analysis of items and instruments

Test characteristics (discrimination D, item-test correlation r_{it}, and internal consistency, calculated as Cronbach's alpha α_C) were calculated and item and instrument analysis were conducted out according to standard procedures and requirements [97,98]. We also calculated, for each item α^*, the value of α if the item is removed, which can be used for scale improvement procedures [99].

For simplified comparison and interpretation, we follow Cohen *et al.* [100], who suggest a linear transformation of raw scores to the fraction (or percentage) of the possible maximum score. The values of psychometric scales for motivational variables and for the conceptual and spatial abilities tests were thus transformed to the interval between 0 and 1, according to $(S_o - S_{min})/(S_{max} - S_{min})$, where S_o is a given observed score, and S_{min} and S_{max} are the minimum and maximum possible scores for a given scale, respectively.

Learning mechanics includes several separate but related concepts, such as interpreting one-dimensional or two-dimensional graphs, understanding subjects including Newton's laws or free fall. In order to test for potential subdimensions of the test of conceptual understanding, a factor analysis was additionally carried out (Ref. [101]; details of the procedure are reported in the Appendix A).

2. Analysis of motivational and learning effects

In order to test for effects of the intervention on motivational and learning outcomes in presence of predictors, analyses of (co)variance (AN(C)OVAs) were carried out [102,103]. First, it was ensured that the underlying assumptions are respected (independence and Gaussian distribution of residuals; homogeneity of residuals' variances; and, for analysis of covariance, homogeneity of regression slopes [103]). Second, in order to test for possible influences of predictors, we follow the approach recommended in Refs. [101,104], combining an *a priori* selection of predictors available in the literature (see Sec. III D 3) with a preliminary statistical test for potentially significant influences (ANOVA for a given combination of outcome and predictor variable). Finally, once all significant predictors covariates were determined, we took them into account in the model and proceeded with the main analysis ANCOVA. A correction for false discoveries was applied using the Benjamini-Hochberg procedure for multiple comparisons [105].

For comparing changes before and after the instruction or between the treatment group and the control group, effect sizes are reported as Cohen's d [106]. Usual effect-size levels (as established from comparison of a great many studies in different areas) are small ($0.2 < d < 0.5$), medium ($0.5 \leq d < 0.8$), and large ($0.8 \leq d$) [106]. Another reference value of $d = 0.4$ is used by Hattie [107] as a "hinge point" between influences of smaller and larger size.[6] Effect sizes when comparing more than two groups (after an ANCOVA) are reported as total eta squared, η_t^2 [106]. Effect-size levels for η_t^2 are small ($0.01 \leq \eta_t^2 < 0.06$), medium ($0.06 \leq \eta_t^2 < 0.14$), and large ($0.14 \leq \eta_t^2$); the same caveat6 as for the d thresholds applies.

Analyses were carried out using the R software package [108] and Excel® and the data are available in Ref. [73].

IV. RESULTS I: INSTRUMENT VALIDATION

In this section, we present an overview of the main results of the item and test analysis for learning and motivational outcomes, as well as for predictor variables. All results refer to the post-test of the main study, if not stated otherwise. Detailed results of the items and instruments of all tests (pilot study, pretests) can be found in Ref. [73].

A. Learning test

The newly developed conceptual learning questionnaire, adapted to the specific learning content of the intervention and the local curriculum (see Sec. III D 1), was analyzed according to standard procedures and requirements (Ref. [97], p. 16; see Sec. III E 1).

The conceptual test was constructed based on several *a priori* hypothesized, potentially interrelated conceptual dimensions: one-dimensional kinematics; free fall; Newton 1st, 2nd, and 3rd laws. The interdependence and, at the same time, the diversity of the items of the conceptual test were consequences of the educational objectives discussed in Sec. III B and is a current property of classroom tests [109,110]. In order to look for potential sub-dimensions of the test, an exploratory factor analysis was carried out. It revealed only one factor with 12 items, related to rectilinear motion, and no other factors composed of more than 3 items. Table IV shows the results of the item analysis for both the full and restricted test (12 and 19 items, respectively).

A comment on the values of α_C is in order: as emphasized by current practice and recommendations, assessment tools covering multiple concepts, especially in classroom settings, are expected to have low values of α, and that low internal consistency is not a significant barrier to the use of a test with meaningful content coverage [102,111]. In their review on assessment instruments in science education, Adams and Wieman [109] also indicate that instruments designed to measure multiple concepts in a short time, such as those administered in typical classroom settings, may have low values of internal consistency. According to van Blerkom [112], typical classroom tests indeed show a range of values between 0.60 and 0.80. Therefore, a value around 0.6 falls in the acceptable range for group level measurements [113,114]. This reasoning is consistent with the results for the single dimension revealed by factor analysis, related to the conceptual understanding of rectilinear motion. As Table IV shows, the increased uniformity of the restricted tests lead to a larger internal consistency (as well as a larger item-test correlation).

In view of the above, we kept the full test (19 items) for further analysis as a reasonable compromise between content coverage and psychometric quality.

B. Motivational variables

As in the case of the learning test, instruments for the four motivational variables (interest, perceived relation to reality, self-concept, and curiosity state) were analyzed according to standard procedures and requirements [97,115]. All test properties were within the recommended ranges, attaining satisfactory to good values [97]; see Table V for post-test values (values at pretest are very similar).

C. Predictor variables

As for the dependent variables, the predictor variable instruments were also analyzed according to standard procedures and requirements [97,115]. Results are reported in Table V and show satisfactory values.

[6]We agree with Hattie [107] that these threshold values are an element of discussion to be used with circumspection, not a criterion to be applied blindly. Lower effect sizes might well be worth considering, depending on available alternatives, effort, and so on, and vice versa for higher effect sizes.

TABLE IV. Instrument analysis of the conceptual post-test of the main study for the full version (19 items) and for the restricted version (12 items) with $N = 111$; M = mean; SD = standard deviation; CI = confidence interval, for α_C.

Characteristics	Full version M(SD/CI)	Restricted version M (SD)	Restricted version Range
Item difficulty P	0.43 (0.19)	0.49 (0.17)	[0.20; 0.76]
Item discrimination D	0.41 (0.22)	0.56 (0.13)	[0.37; 0.75]
Item-test correlation r_{it}	0.34 (0.14)	0.45 (0.08)	[0.35; 0.57]
Internal consistency α_C	0.59 (0.11)	0.68 (0.09)	⋯

TABLE V. Sample items and instrument analysis of the motivational variables [*interest* (IN), perceived *relation to reality* (RR), *self-concept* (SC), and *curiosity state* (CS)] at post-test (α_C: Cronbach alpha; r_{it}: item-test correlation; SD: standard deviation; and CI: confidence interval).

Variable: example of item (translated from French)	α_C (CI)	Mean r_{it} (SD)	Range of values r_{it}
IN: I liked the physics class.	0.71 (0.09)	0.69 (0.25)	[0.56; 0.76]
RR: The subjects of the physics course are useful for everyday situations.	0.90 (0.03)	0.85 (0.15)	[0.79; 0.90]
SC: My classmates thought I was good at physics.	0.89 (0.03)	0.84 (0.19)	[0.75; 0.89]
CS: The course aroused my curiosity about the topics covered.	0.89 (0.03)	0.83 (0.12)	[0.81; 0.86]

V. RESULTS II: INTERVENTION EFFECTS

In this section, we present the main results obtained on the whole sample participating in the main study.

A. Learning

Table VI reports descriptive values of the concept test before and after the instruction, specified for class groups (G1–G4: pairs of class groups taught by different teachers), the control group and the treatment group, as well as for all participants, together with values for Cohen's d. The results of the paired t tests revealed significant gains for learning, with large effect sizes ($d > 0.8$) for all but one group (we will come back to the one group with a medium-sized effect below). Thus, the instruction was globally effective, both for the treatment group and the control group (it successfully removed most of the preexisting misconceptions: the number of the most frequent incorrect answers fell from 60% in the pretest to 25% in the post-test).

Furthermore, the following findings of the ANCOVA about learning outcomes are noteworthy (see Table VI). No significant effect was observed of the independent variable *treatment* ($F(1, 109) = 0.026$, $p = 0.87$). A medium-sized effect of the variable *class group* (G1–G4) on the *learning post-test* (LPO) results was observed [$F(3, 96) = 5.2$, $p < 0.01$, $\eta_t^2 = 0.10$], as well as a small effect of the predictor *learning pretest* [$F(1, 96) = 8.6$, $p < 0.01$, $\eta_t^2 = 0.06$]. No other predictors turned out to be significant for the conceptual post-test, specifically, no significant effects were found for *gender*, cognitive variables (prior *grades in mathematics, physics,* and *French, spatial abilities,*

TABLE VI. Descriptive values for *learning* pre and post-test (LPR, LPO) and t-test statistics for the different groups : control group (CG) and treatment group (TG); G1–G4: pairs of class groups taught by different teachers): mean success rates P ("difficulties," see Ref. [97]), standard deviations, t-values, p-values, and Cohen's d for pre- and post-test gains.

Group	N	P (SD) LPR	P (SD) LPO	t	p	Cohen's d
G1	27	0.28 (0.15)	0.49 (0.14)	5.03	<0.001	1.4
G2	26	0.27 (0.08)	0.43 (0.16)	4.70	<0.001	1.3
G3	27	0.25 (0.11)	0.32 (0.16)	1.91	<0.001	0.5
G4	31	0.36 (0.14)	0.49 (0.14)	3.78	<0.001	0.9
TG	55	0.31 (0.13)	0.44 (0.17)	4.26	<0.001	0.9
CG	56	0.27 (0.13)	0.43 (0.15)	5.95	<0.001	1.1
Tot	111	0.29 (0.13)	0.43 (0.16)	7.17	<0.001	1.0

TABLE VII. Significant results of the ANCOVA analysis on the concept learning post-test (LPO: learning post-test; class group; LPR: learning pretest). The symbols * and ** indicate small and medium effects, respectively.

LPO	Sum squared	Degrees of freedom	F-value	p	η_t^2
Treatment	7.45	1	1.18	0.28	⋯
Class group (G1–G4)	101.1	3	5.34	<0.01	0.10**
LPR	59.7	1	9.45	<0.01	0.06*
Residuals	600.0	95			

TABLE VIII. Mean value [total/treatment group (TG)/control group (CG)] and standard deviations of *cognitive load*, *cognitive activation*, and motivational variables (CT: *curiosity trait*; CLE and CLS: *cognitive load related to the experiments* and to the use of *the smartphone app* respectively; CAE: *cognitive activation during the experiments*; INV: student *involvement* during the physics lessons; SCS: *self-concept regarding use of smartphones* and *apps*; TA: *teacher assessment*).

Variable and sample item (translated from French)	Mean value (SD): total/TG/CG	Range
CT: I find it fascinating to learn new things.	4.62(0.71)/4.72(0.73)/4.53(0.69)	[3.79; 5.22]
CAE: I was actively involved in doing the experiments.	4.16(0.73)/4.07(0.73)/4.24(0.74)	[3.84; 4.59]
INV: I shared my thoughts during the class discussions.	3.60(0.95)/3.56(0.93)/3.63(0.99)	[3.19; 3.99]
SCS: I am good using the apps for smartphone.	4.48(0.88)/4.40(0.84)/4.55(0.92)	[4.19; 4.79]
CLE: I could well focus on the experiments, without struggling with the equipment.	4.06(0.82)/4.10(0.87)/4.03(0.78)	[3.40; 4.68]
CLS: I could well focus on the experiments, without struggling with the apps for smartphone.	Only TG: 4.35 (1.15)	[3.78; 4.76]
TA: The teacher's explanations helped me to understand.	4.91(0.43)/4.88(0.56)/4.93(0.27)	[4.46; 5.24]

cognitive load and *activation* during the experiments, *cognitive load* using the apps, *involvement* during the lessons), nor affective variables (*relation to reality, self-concept* regarding physics or using the apps, and *curiosity state* and *trait*).[7]

An additional ANCOVA on the individual item level of the learning post-test was carried out. Taking into account the correction for multiple significance testing [105], no differences were found between the control and the treatment groups.

Moreover, an ANCOVA for the *post average grade in mathematics* was carried out, and we found an indication of a possible positive effect of the treatment [$F(1, 102) = 2.9$, $p = 0.09$, $\eta_t^2 = 0.03$]. Although the p value for this hypothetical effect is only marginally significant, it is worth considering the possibility of such an effect, as we will discuss in Sec. VII.

B. Cognitive load, cognitive activation, and motivational variables

First, the mean values of the motivational and other control variables for all participants, as well as for the control group and the treatment group are reported in Table VIII. We observe that the values for all these variables are comparable for the control group and for the treatment group. In particular, there was no difference in *cognitive load during the experiment* between the learning groups with and without MDETs [$t(107) = 0.48$; $p = 0.6$].

Second, descriptive values for the motivational dependent variables (*interest, perceived relation to reality, self-concept,* and *curiosity state*) before and after the instruction are reported in Table IX. The t test results showed for the whole sample a significant gain for *perceived relation to reality* [RR, $t(219.57) = 2.2$, $p = 0.03$], with an effect size of $d = 0.29$. No significant changes were found for *interest* [IN, $t(219.93) = 1.6$, $p = 0.12$], *self-concept* [SC, $t(216.6) = 0.28$, $p = 0.77$], and *curiosity state* [CS, $t(215.87) = 0.60$, $p = 0.55$].

From the ANCOVA results, no significant effects of the treatment (control group or treatment group) on the motivational dependent variables were found [$F(1, 109) = 0.92$, $p = 0.34$ for *interest*; $F(1, 109) = 0.65$, $p = 0.42$ for *curiosity state*; $F(1, 109) = 0.009$, $p = 0.92$ for *self-concept*; and $F(1, 109) = 0.27$, $p = 0.61$ for *relation to reality*].

Table X reports the ANCOVA results for motivational outcomes, focusing on those variables, where significant pre- and post-test changes have been found.

Interest: A medium effect of the variable *class group* (G1–G4) was found in the post-test, even when controlling for *involvement* and *cognitive activation* during the experiments [$F(3, 101) = 3.3$, $p = 0.02$, $\eta_t^2 = 0.04$, see

[7]As the predictor *teacher assessment* is highly correlated with the variable *class group*, we did not take it into account for the ANCOVA.

TABLE IX. Descriptive values and t test statistics of the pre- and post-tests for motivational variables: mean values (M), standard deviation (SD), and t- and p-values (6-level scale, 1 = completely disagree, 6 = completely agree), for TG (treatment group)/CG (control group)/total.

		Interest (IN)						Curiosity state (CS)			
		M (SD)	M (SD)					M (SD)	M (SD)		
Group	N	Pre	Post	t	p	Group	N	Pre	Post	t	p
TG	55	3.10 (0.88)	3.27 (0.90)	1.0	0.31	TG	55	3.40 (1.04)	3.34 (1.05)	0.29	0.77
CG	56	3.24 (0.82)	3.43 (0.83)	1.2	0.23	CG	56	3.28 (0.85)	3.17 (1.13)	0.56	0.57
Tot	111	3.17 (0.85)	3.35 (0.87)	1.6	0.12	Tot	111	3.34 (0.95)	3.25 (1.09)	0.60	0.55
		Self-concept (SC)						Perceived relation to reality (RR)			
		M (SD)	M (SD)					M (SD)	M (SD)		
Group	N	Pre	Post	t	p	Group	N	Pre	Post	t	p
TG	55	3.52 (1.07)	3.59 (1.17)	0.31	0.76	TG	55	3.44 (1.04)	3.73 (1.00)	1.4	0.15
CG	56	3.59 (0.97)	3.61 (1.15)	0.09	0.93	CG	56	3.25 (1.12)	3.61 (1.25)	1.6	0.11
Tot	111	3.56 (1.02)	3.60 (1.15)	0.28	0.77	Tot	111	3.35 (1.08)	3.67 (1.13)	2.2	0.03

Table X]. Note that no other predictors turned out to be significant for interest at post-test.

Self-concept: Significant effects (with small effect sizes) of the variables prior *self-concept* [$F(1,89) = 3.3$, $p < 0.01$, $\eta_t^2 = 0.05$], and *physics grade* [$F(1,89) = 3.1$, $p < 0.01$, $\eta_t^2 = 0.04$], as well as of *gender* on *self-concept* were found, consistent with prior research [116–118]. No other predictors turned out to be significant for self-concept at post-test.

Curiosity state: There were significant small-to-medium effects of the predictors *prior curiosity state* [$F(1,99) = 18.6$, $p < 0.01$, $\eta_t^2 = 0.09$], *cognitive activation* during the experiments [$F(1,99) = 4.6$, $p = 0.04$, $\eta_t^2 = 0.02$], and *involvement* [$F(1,99) = 4.2$, $p = 0.04$, $\eta_t^2 = 0.02$], on the

TABLE X. ANCOVA results for motivational outcome variables at post-test; * and ** indicate small and medium effects, respectively. Outcome variables are POST_IN (*interest* at post-test), POST_SC (*self-concept* at post-test), POST_CS (*curiosity state* at post-test), POST_RR (*perceived relation to reality* at post-test. Predictors include CAE (*cognitive activation related to the experiments*), CLE (*cognitive load related to the experiments*), INV (student *involvement* during the physics lessons), PRE_CS (*curiosity state* at pretest), PRE_SC (self-concept at pretest), PRE_RR (perceived relation to reality at pretest) and PY_GRADE_POST (*physics average grade* at post-test)

POST_IN	Sum of squares	Degrees of freedom	F value	p	η_t^2
Class group (G1–G4)	3.5	3	3.3	0.02	0.04*
CAE	2.1	1	6.1	0.02	0.06**
INV	5.1	1	14.4	<0.01	0.12**
Residuals	35.6	101			
POST_SC					
PRE_SC	3.3	1	9.3	<0.01	0.05*
GENDER	2.4	1	6.8	0.01	0.03*
PY_GRADE_POST	3.1	1	8.6	<0.01	0.04*
CLE	1.9	1	5.3	0.03	0.03*
Residuals	14.0	89			
POST_CS					
PRE_CS	12.4	1	18.6	<0.01	0.09**
CAE	3.0	1	4.6	0.04	0.02*
INV	2.8	1	4.2	0.04	0.02*
Residuals	65.9	99			
POST_RR					
PRE_RR	11.7	1	14.0	<0.01	0.08**
CAE	2.5	1	3.0	0.09	n.s.
CLE	2.5	1	3.0	0.09	n.s.
Residuals	84.1	101			

variable *curiosity state* at post-test. No other predictors were found to be significant for self-curiosity at post-test.

Perceived relation to reality (POST_RR): Only the pretest value (PRE_RR) was found to have an effect on the post-test *perceived relation to reality* [medium effect size, $F(1, 101) = 14.0,\ p < 0.01,\ \eta_t{}^2 = 0.08$].

VI. DISCUSSION

A. Cognitive aspects

Our findings show that an overall learning effect occurred, with a pre- and post-test effect size (Cohen d) around 1, indicating a large effect. The beneficial effect of the instruction on conceptual learning also occurred separately in all individual class groups. However, although an overall strong learning gain of conceptual understanding of mechanics was observed both for the control group and the treatment group, the answer to our first research question is that there is no significant difference in physics learning between the control group and the treatment group. This result differs from those of the studies by Hochberg *et al.* [31] and Klein *et al.* [27], discussed above. A possible reason could be the differences in the target groups of these studies. Here we examined high school learners of non-specialized physics classes, while the learners for whom the positive effects had been found in Hochberg *et al.* [31] were from high school specialized physics classes, and in Klein *et al.* [27] were first year physics university students. In both cases, these are learners with advanced physics understanding, and this might be a prerequisite for fully exploiting the cognitive potential of MDETs. Another factor for further improvement may be due to the setting of regular mechanics course with its tight constraints, limiting feasible experiments and topical contexts (see Sec. III C).

On the other hand, similarly to what found by Hochberg *et al.* [19,31] and Klein *et al.* [27], the use of MDETs did not increase the perceived cognitive load during the experiments. Moreover, the ANCOVA revealed no significant influence of perceived cognitive load due to the use of apps on the learning outcomes of the treatment group. Thus, no evidence of often apprehended disadvantages of over challenging and distraction of learners by mobile devices was found in the present study.

Regarding features of successful labwork, another point beyond the use of MDETs is worth discussing. Recent research shows that traditional, highly structured undergraduate labs, reinforcing textbook content (e.g., verifying Hooke's law), do not enhance learning [119,120]. Smith *et al.* [119,120] define such labs as combining content verification with rigid "cookbook" formats, leaving little room for autonomous student thought. We fully agree these formats are ineffective. While our lab course includes a verification role focused on conceptual obstacles, its goals align well with those of the authors cited above (e.g., Ref. [121]). Specifically, our labs provide procedural guidance (though no "cookbook" protocols), aiming for content enhancement rather than mere reinforcement. They allow students to confront conceptual obstacles with hands-on, empirical evidence—a well-founded purpose of student experiments, especially at the high school level. With this clarification—that verification (or "enhancement") does not imply cookbook labs—we see a shared key factor with [119,120]: proper "active learning engagement" [122,123] or "cognitive activation" [124,125] for the given learning goals (in our case conceptual understanding). Our results confirm that MDETs labs can indeed provide such activation.

An effect of the different class groups (different classes taught by a different teacher) on the result of the conceptual test was observed (see Sec. V A and Table VII), with a medium effect size. This effect, mainly due to the lower results of group 3 (see Table VI), can be explained by the fact that the learners of this group completed the post-tests during the first COVID-19 lockdown as a formative assessment, without any influence on the final grade. This may have affected the learners' investment in learning both before and during the conceptual post-test.

A final observation is worth mentioning. While previous studies generally highlight the positive impact of mathematical understanding on physics learning [126–131], our study suggests a possible reverse effect of using mathematics in a physics course context: the use of MDETs and of the cognitive effort required to understand multiple representations generated by them may enhance mathematics learning (see, e.g., Refs. [132,133] for similar considerations).

B. Motivational aspects

Overall, the impact of the instruction was similar for the control group and the treatment group concerning motivational aspects. The pre- and post-test variations of the perceived *relation to reality, self-concept*, *interest*, and *curiosity state* in physics are similar for both groups and, in addition, they were all slightly positive. For *perceived relation to reality*, this improvement was even significant for both the treatment and control groups, with a small effect size ($d = 0.29$). These results show that it was possible to maintain or even slightly improve the motivational aspects investigated in this study. Regarding the comparison between the control group and the treatment group, no significant differences were found. Thus, regarding our second research question on motivational aspects, a slight improvement could be established, but no superior effects by the MDETs treatment. In addition, we collected the participating teachers' perceptions on using mobile devices in physics labs, and none reported the observation of over challenging or distracting effects on students' focus. They emphasized the practicality and ease of MDETs, which not only reduced repetitive tasks but also enabled experiments that would not be possible with traditional equipment. While their attitudes ranged from cautious to

optimistic, all agreed on the value of the approach and expressed their intention to continue using MDETs in future classes.

C. Limitations

The following methodological limitations of the study have to be noted. First, there are interesting variations of the use of MDETs which could not be integrated in the design. One such possibility is using the learners' private smartphones for the MDET activities Bring Your Own Device (BYOD; [134,135]), which was not explored in this study, as in the Geneva schools the use of private smartphones during the lessons is prohibited. Moreover, this idea raises several additional issues, which would have strongly changed the intentions of this research. First, from a learning point of view, the distracting effects by private smartphones are likely to be stronger than those of devices provided by the institution, as the possibilities of personal use (images, music, games, and communication) are present. Second, although smartphones and tablets are very widespread, some learners can afford models that perform better than others and some learners do not have any at all; therefore, the question of equal treatment should also be considered. Furthermore, the apps used in the study are only available for a brand of high-end devices, and there are no similar apps for all models of tablets or smartphones. Asking the families of all learners to buy apps for their phone is problematic: families can be opposed, many learners do not have the right brand and, to date, no similar free-to-access video tracking apps are available. Thus, while the BYOD approach is advocated by some authors, and indeed seems to have also promising features [136–138], it was not included in this study. A further discussion of pros and cons can be found in Ref. [139].

Second, there are limitations on the level of the variables taken into account in the present study. One such limitation is about cognitive load and the differentiation of its different types. For instance, high values of CL can be either positive or negative for learning, depending on the proportion of extraneous and germane cognitive load [140,141]. Possibilities for empirically differentiating between the different types of CL exist [142], but were beyond the scope of this investigation.

Finally, the variable "group" (i.e., a certain class taught by a certain teacher) has a considerable influence on the outcomes for learning and interest. But the design and instruments of the present study do not allow for a more fine-graded analysis of the potential factors at work. This is a natural limitation of quasiexperimental investigations in real classroom settings, with their tight constraints, e.g., on available testing time, and thus the number of variables which can be tested. Note, however, that in the present study two measures to control for possible teacher influences were taken: the same teacher taught both a control and a treatment class, and additionally, one dimension of the WIHIC classroom questionnaire [102] was taken into account, viz., teacher engagement (as perceived by students). While this allows for some control of overall (same teacher) and differential teacher effects (engagement questionnaire) on the comparison of treatments, a more detailed investigation would be of interest (see below).

A limitation of the present study is its setting within a regular mechanics course, which restricts the types of student experiments that can be implemented, thereby not fully exploiting the potential of topical contexts. Nevertheless, we see our work as a valuable contribution to expanding the research foundation for integrating MDETs into everyday instruction. Note also that, despite the above choices and their acknowledged limitations, the MDETs intervention still leads to strong pre- and post-test gains in conceptual learning and even some gains in motivational outcomes. These findings, in turn, can serve as a basis for developing and studying more far-reaching ways of creating contexts with MDETs in more open teaching settings.

Finally, the following practical limitations should be considered when teaching with MDETs with high school students: first, there is a phase of appropriation of the mobile device (tablet or smartphone) and of the apps. It is generally not easy for learners to make a video of sufficient quality for the tracing by the app to be possible: it has to be stable during the recording, sufficient contrast between the studied object and the background is needed, the angular speed (pixel speed) cannot be too high, the tracing cursor has to be chosen with an appropriate width. All these requirements take time to be learned and mastered, which may not be evident for learners and, often, to teachers as well (one among the four participating teachers had expressed a certain degree of difficulty in that respect). Moreover, it is worth noticing that certain teachers who did not participate in the study did not even take into consideration to use these devices in their lessons, because they mainly affirmed that they could not master them with sufficient ease. This confirms what has emerged in previous investigations on the difficulties of integrating technologies in schools, including human factors or other barriers due to the technologies [143].

Second, beyond the practical mastery of the device, the use of MDETs may cause cognitive overload (see Sec. II C), if learners do not yet master the concepts and mathematics underlying the representations provided by the apps and used during the experiments (e.g., diagrams of position or of velocity vs time). On the one hand, it is well known that familiarizing with the different mathematical representations of a physical phenomenon requires an effort from learners [64,140]. On the other hand, it might be this effort that could give a better knowledge of mathematics, see below. Thus, when planning a teaching sequence, teachers should carefully try to maintain a balance between productive efforts for learning and the risk of overloading the working memory. The present study shows that it is

possible to do so, but teachers should be aware that this requires special attention.

Third, despite initial intentions, it was not possible to let the learners carry out MDET activities as homework, mainly for practical reasons. They would have had to take the devices home, which was not manageable from an organizational point of view for the limited availability of these devices (several classes in parallel used the same set of tablets furnished) and for the security of their use.

Finally, difficulties may arise in the support of tablets by schools, taking into account the maintenance, and the technical assistance they request. It may seem trivial, but, for example, having a Wi-Fi network or a Bluetooth printer in each classroom is not obvious. Updates are constantly needed, and these high-tech objects become obsolete in a few years, faster than standard lab equipment like rails, timers, or dynamometers. Furthermore, we cannot neglect the planned obsolescence of these devices and their cost in environmental and durability terms [144]. All these factors should be put on a balance with hindsight, taking into account the global context and the fact that we imagine a scientific education sustainable and accessible to all the community.

VII. PERSPECTIVES AND CONCLUSIONS

A. Perspectives for classroom practice and further research

We begin with a series of observations regarding classroom practice. Taking into account its beneficial effects as well as the limitations described above, the use of MDETs turns out to be advisable under certain conditions, not as a replacement, but as a complement of traditional settings. For some course contents, it is possible to choose whether to do an experiment with traditional instruments or with MDETs (as for linear motions), while for other contents either the traditional activity (e.g., free fall, where the temporal resolution of the MDETs app is insufficient) or the MDETs activity (e.g., the "jump" activity, see Sec. III C and Ref. [4]) are clearly more advisable.

When possible, the easiness and practicality of the preparation of the experiments by the teacher constitute a clear advantage: for the same content as in a conventional activity, MDET activities allow faster implementation with a handy and light device, allowing to save precious teaching time for other objectives. This advantage was highly appreciated by teachers participating in this study.

Teachers also appreciated the flexibility of MDET in different classroom settings: depending on the class group and the lesson plan, learners can take measurements and analyze them (experimental activity), or the data can be given from the measurement by a single student or by the teacher and used by the whole class (semiexperimental activity). In the latter case, the data ownership and the real-time data capture are not present, but the possibility of learning in authentic situations might be even enhanced, as the video may be recorded outside the classroom (e.g., playground, amusement park).

The following perspectives appear as interesting for both research and practice. First, the possibility of "home experiments." Learners can carry out activities at home, for example, as homework or when the lessons are given at a distance. This possibility has been considered in science education for almost two decades [145], and it has attracted new attention during the COVID-19 period and beyond [146]. Educational institutions worldwide were obliged to adapt to remote learning [147,148], and MDETs have shown to be a promising element in these strategies [149]. The insights learned from these experiences continue to improve physics instruction. An empirical investigation of these ideas was not in the scope but would be of interest both for practice and research in science education.

Second, we mention the potential of MDET activities to create a closer interaction between physics and mathematics education [150]. For example, mathematics classes could use representations obtained from physics experiments to illustrate how mathematical concepts apply to real-world contexts. Conversely, in physics courses, MDET can support an enhanced integration of mathematics. An empirical study on the effects of MDET on students' learning of mathematics as applied to physics thus appears promising.

Finally, the following aspects are also of interest for further research. We already mentioned that the differentiation of different types of cognitive load [142], potentially coupled to measures of short-term memory capacity, appears to be relevant for future research on the implications of the use of MDETs in the learning of physics. Another topic worth exploring is how to improve the understanding of the "group" factor (a given combination of teacher and class) in terms of operationally measurable variables. Such variables could be assessed at a general level, for example, using one of the instruments of the WIHIC questionnaire (What is Happening in this Class [102,151,152]), such as the dimensions of "investigation" and "task orientation." On a more specific level, attitudes and competences of teachers regarding the use of technology in the science classroom are of interest, such as analyzed in the TPCK framework (technological pedagogical content knowledge [153]).

B. Summary and conclusions

In this research, we examined the impact of integrating mobile devices as experimental tools (MDETs) in physics laboratory sessions over one semester. In particular, we investigated how this integration influenced situational interest, curiosity, relation to reality, self-concept, conceptual learning in physics, and the acquisition of related mathematics among students enrolled in a nonspecialized mechanics course. Among our findings, we highlight the following main points.

First, this is—to the best of our knowledge—the first empirical study about the educational effects of MDETs in a full teaching sequence on physics at the high school level. Specifically, the present contribution presents such an investigation in a mechanics high school course, with a quasiexperimental pre- and post-test design comparing a treatment group with MDET vs a control group without MDET, and taking into account various control measures (same content, lesson plan, and teacher for control and treatment group; use of series of predictor variables). For the global learning gain, a statistically significant effect with an effect size $d = 0.9$ was found, which is quite a large effect. Compared to the control group, no significant difference was found. This appears unexpected in view of the theoretical advantages of the approach (Sec. II); however, as the learning gains of both control and treatment groups are quite large, one is dealing with a quite effective teaching in both cases. Possible reasons for the absence of a greater learning gain of the MDET group are discussed. Moreover, evidence for "null effects" for new treatments, especially when accompanied with a certain "hype" of expectations as it is often the case with instructional technology, is considered as increasingly important in educational, social, health, and other sciences [154,155], including physics education research [156].

Second, the present study did not find evidence supporting commonly anticipated drawbacks of instructional technology. Specifically, there was no difference regarding cognitive load between the control group and the treatment group. This suggests that when the use of mobile digital educational technologies (MDET) is appropriately planned and guided by a teacher, they do not over-challenge students. Moreover, the presence of positive learning gains and the absence of negative effects did not depend on various potential influence factors; in particular there was no or little dependence on gender, prior knowledge, and other predictors. Put positively, the findings of this study can be summarized by saying that students can learn by the MDETs approach as much as by well-prepared highly effective conventional lessons (but not more), and that learners of both genders and with different initial conditions can have comparable learning gains (of course, the absolute level depends on the initial state).

Finally, MDETs thus appear as a useful extension of the toolkit of the physics teacher, comparable to other existing successful teaching practices, while offering a number of additional advantages (e.g., variation of teaching approaches, inclusion of digital skills) and potential further perspectives for research and development in physics education (extensions to nonstandard physics experiments, e.g., as homework; synergies between physics and mathematics learning).

ACKNOWLEDGMENTS

Swiss universities has supported this project through its program on discipline-based education research. The implementation of the teaching materials and the study as a whole was only possible thanks to the close collaboration and very good mutual understanding between researchers, teachers, and school laboratory assistants throughout the process. Moreover, valuable technical support by Dr. L. Darmendrail (now Valdivia, Chile) and helpful advice by Dr. L. Weiss (Geneva) are gratefully acknowledged.

Conceptualization: A. G, A. M; Formal Analysis: A. G., M. D., F. S.; Investigation: A. G, A. M; Methodology: A. G., M. D, A. M., F. S.; Project Administration: A. G, A. M; Supervision: A. M.; Writing: A. G, A. M, with contribution by M. D., F. S.

DATA AVAILABILITY

The data that support the findings of this article are openly available [73].

APPENDIX A: FACTOR ANALYSIS

Exploratory factor analysis was performed for the conceptual test, in view of the complexity of the concepts' interdependence in the targeted physics curriculum, and the absence of previous empirical evidence. The number of factors was determined by considering only the factors with eigenvalues >1 [101]. The resulting subject-to-item ratio was about $6:1$ ($N = 111$, $k = 19$ items), which can be considered as sufficient [157]. In order to identify the most suitable factor structure, we compared several solutions with different numbers of factors, used oblique rotation to simplify the data [157], and selected the solution that best aligned with theoretical and/or educational considerations based on a series of "goodness of fit" indices [158,159]. Following [160], we considered that an item belonged to a given factor when its loading was greater than 0.3, unless a theoretical argument supported its presence.

APPENDIX B: EXAMPLES OF MDET ACTIVITIES

The round trip: This activity consists of analyzing a return trip movement of a student (see Fig. 2). Each trip covers the same distance but is made at a faster speed on the outward journey than on the return. The learners have to calculate the velocity vector and the speed for the outward journey, for the return, and for the entire trip, after having made their prediction of the total average speed. This activity therefore allows the learners to confront the misconception that the total average speed is equal to the mathematical average of the speeds in each of the journeys of the same distance: according to the studies of Reed *et al.* [161,162], 85% of the students give this answer, and only 5% of the students give the right answer. For this activity, it was possible to keep the same related-to-reality situation for the control group and the treatment group: the learners of the control group could do the exact same

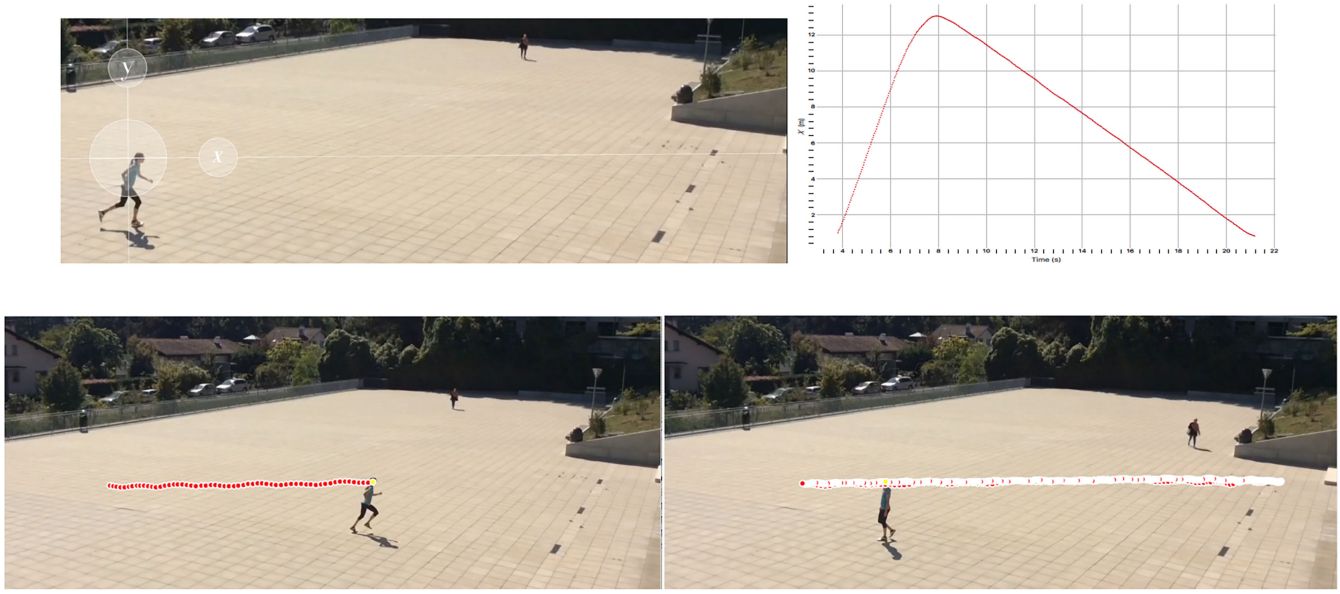


FIG. 2. Example of a situation of round trip MDETs activity: fixing the reference frame, time diagram, and tracing the movement.

activity, but measuring times and distances with traditional instruments, such as stopwatches or meters.

Uniform linear motion (ULM): To analyze ULM, control group learners performed an experiment using a traditional setup, which varied according to the teacher and the school: for example, the motion of a ball over a short distance, of a marble or a cart on a rail (air rail or slightly inclined rail, to compensate for friction). The learners of the control group had to create the motion's time diagram from the data collected with traditional instruments like a chronometer and a measuring tape. The learners of treatment group analyzed the movement of an overhead projector film roll on a table (see Fig. 3), starting from the data table provided by the app. As for the control group, they calculated the slope and the intercept, and they found that the diagram of velocity is a plateau. Indeed, they also got the graphics directly available by the app and thus did not have to draw it. As one of the pedagogical objectives of this activity was constructing a time diagram of a motion, learners of the treatment group did not use the option "add a regression" to the graph here initially. They were shown this function only at the end of their analysis, to compare with their results for the slope and the intercept.

For both groups, this activity was the first allowing the learners (i) to work on the notion of function and dependent and independent variables as well as (ii) to associate these abstract concepts with the corresponding physical quantities. Although the learners had seen these concepts before in math classes, this topic remains challenging for them.

Uniformly accelerated linear motion (UALM): For the study of this motion type, both treatment and control groups worked with the classroom classical setup (inclined rail, see Fig. 4 on the left) and produced graphs of velocity (linear) and of position (quadratic) vs time for this motion

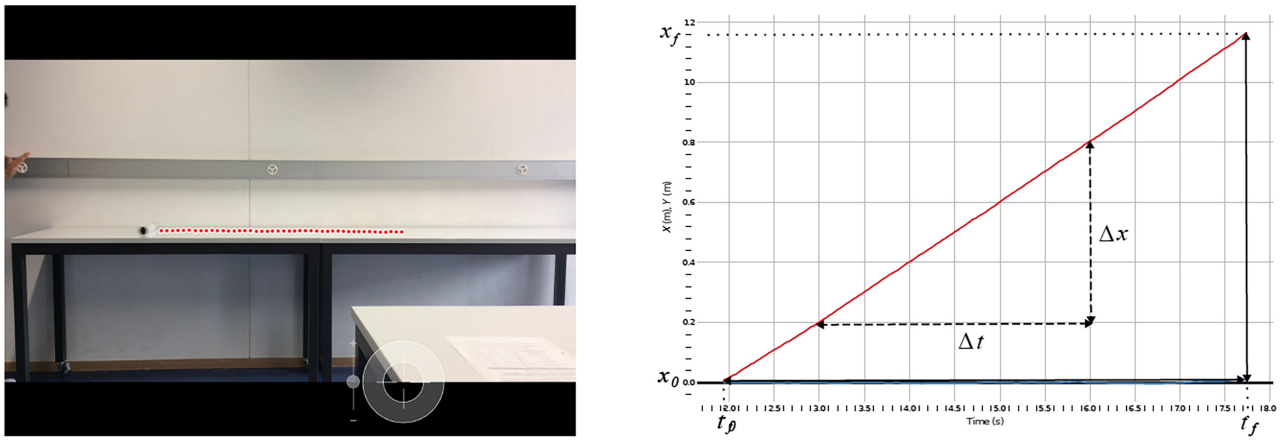


FIG. 3. ULM of an overhead projector film roll: the chronophotography and the time diagram for the same motion.

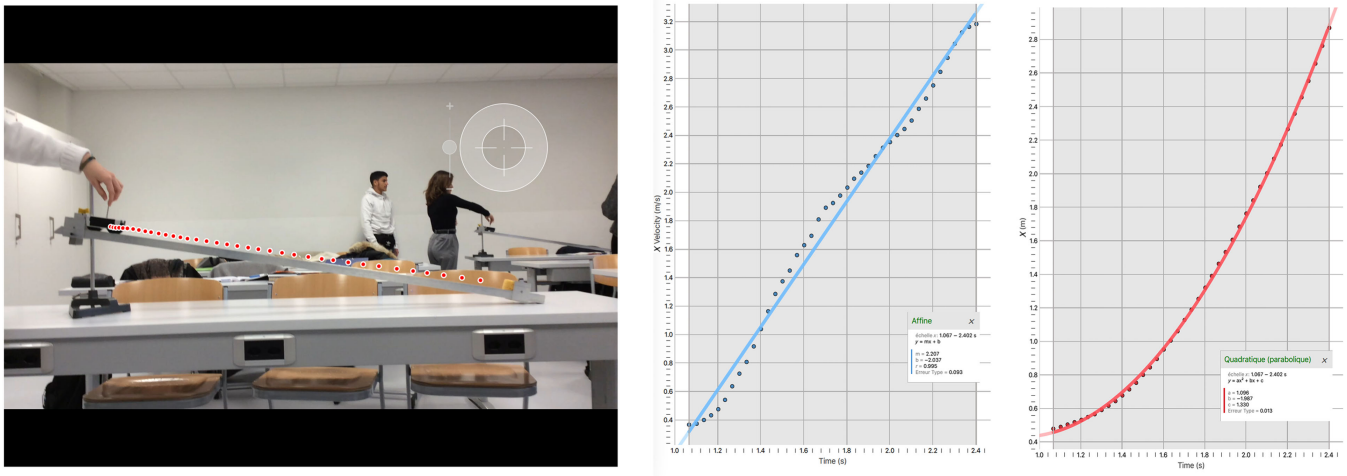


FIG. 4. Example of a MDETs activity on UALM, including chronophotography, position-time and velocity-time graphs.

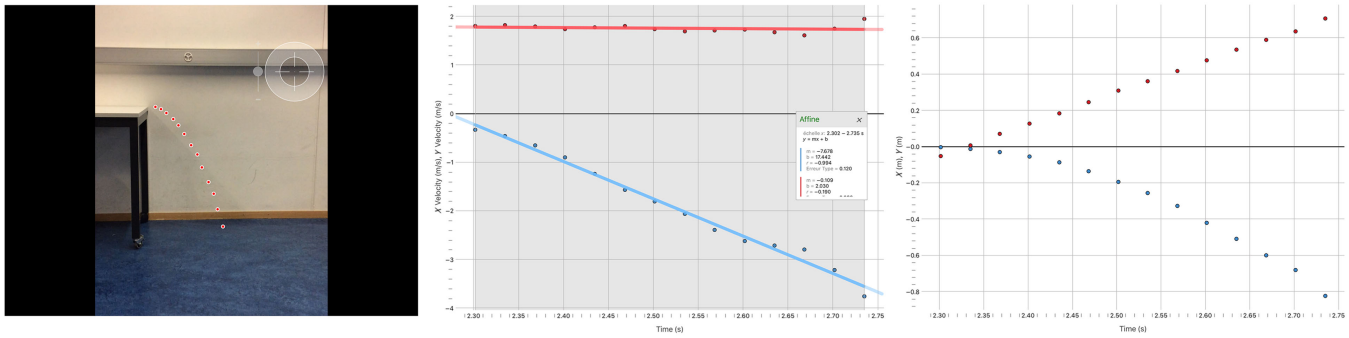


FIG. 5. Tracing, velocity, and position-to-time diagrams outcomes of the apps for the projectile motion activity.

(see the graphs in Fig. 4). As for the ULM activity, the control group learners had to produce the graphics reporting the data, while the treatment group learners used the graphics provided by the apps. Moreover, a regression curve can easily be added, so that the treatment group learners could better focus on the taught physics contents.

Projectile motion: In this activity learners of the treatment group traced the motion of a ball falling to the floor after rolling on a table (i.e., with an initial horizontal velocity). They first performed an analysis of the ULM (x axis) and UALM (y axis) by calculating the velocity or the acceleration from the graph given by the application, as shown in the example in Fig. 5. Subsequently, learners repeated the tracing for different initial speeds. Each time they measured the horizontal distance traveled by the ball during the flight and they established a proportionality relation between this distance and the initial speed. This allowed realizing that the fall time does not depend on the initial velocity but only on the height of the table.

The equivalent traditional lab could be carried out by the control group using a stopwatch and a ruler to measure the ball's initial velocity (from its diameter and passing time) and the horizontal distance from the table when the ball hit the ground.

[1] M. Monteiro and A. C. Martí, Resource letter MDS-1: Mobile devices and sensors for physics teaching, Am. J. Phys. **90**, 328 (2022).

[2] *Smartphones as Mobile Minilabs in Physics: Edited Volume Featuring More than 70 Examples from 10 Years the Physics Teacher-Column IPhysicsLabs*, edited by J. Kuhn and P. Vogt (Springer Nature, Berlin, 2022).

[3] P. Gabriel and U. Backhaus, Kinematics with the assistance of smartphones: Measuring data via GPS – Visualizing data with Google Earth, Phys. Teach. **51**, 246 (2013).

[4] L. Darmendrail and A. Müller, Standing vertical jumps, sport physics with smartphones, Phys. Teach. **59**, 264 (2021).

[5] K. Forinash and R. F. Wisman, Smartphones as portable oscilloscopes for physics labs, Phys. Teach. **50**, 242 (2012).

[6] A. Müller, P. Vogt, J. Kuhn, and M. Müller, Cracking knuckles—A smartphone inquiry on bioacoustics, Phys. Teach. **53**, 307 (2015).

[7] M. Á. González and M. Á. González, Smartphones as experimental tools to measure acoustical, and mechanical properties of vibrating rods, Eur. J. Phys. **37**, 045701 (2016).

[8] A. Müller, M. Hirth, and J. Kuhn, Tunnel pressure waves—A smartphone inquiry on rail travel, Phys. Teach. **54**, 118 (2016).

[9] M. Osorio, C. J. Pereyra, D. L. Gau, and A. Laguarda, Measuring and characterizing beat phenomena with a smartphone, Eur. J. Phys. **39**, 025708 (2018).

[10] M. Monteiro and A. C. Martí, Using smartphones as hydrophones: Two experiments in underwater acoustics, Phys. Educ. **55**, 033013 (2020).

[11] J. A. Sans, J. Gea-Pinal, M. H. Gimenez, A. R. Esteve, J. Solbes, and J. A. Monsoriu, Determining the efficiency of optical sources using a smartphone's ambient light sensor, Eur. J. Phys. **38**, 025301 (2016).

[12] M. C. Sullivan, Using a smartphone camera to explore ray optics beyond the thin lens equation, Am. J. Phys. **90**, 610 (2022).

[13] E. Arribas, I. Escobar, C. P. Suarez, A. Najera, and A. Beléndez, Measurement of the magnetic field of small magnets with a smartphone: A very economical laboratory practice for introductory physics courses, Eur. J. Phys. **36**, 065002 (2015).

[14] M. Monteiro, C. Stari, C. Cabeza, and A. C. Marti, Magnetic field "flyby" measurement using a smartphone's magnetometer and accelerometer simultaneously, Phys. Teach. **55**, 580 (2017).

[15] J. Kuhn, A. Molz, S. Gröber, and J. Frübis, iRadioactivity—Possibilities and limitations for using smartphones and tablet PCs as radioactive counters, Phys. Teach. **52**, 351 (2014).

[16] O. Keller, M. Benoit, A. Müller, and S. Schmeling, Smartphone and tablet-based sensing of environmental radioactivity: Mobile low-cost measurements for monitoring, citizen science, and educational purposes, Sensors **19**, 4264 (2019).

[17] S. Gil and J. Di Laccio, Smartphone una herramienta de laboratorio y aprendizaje: Laboratorios de bajo costo para el aprendizaje de las ciencias, Latin-Am. J. Phys. Educ. **11**, 5 (2017).

[18] C. Stampfer, H. Heinke, and S. Staacks, A lab in the pocket, Nat. Rev. Mater. **5**, 169 (2020).

[19] K. Hochberg, J. Kuhn, and A. Müller, Using smartphones as experimental tools—Effects on interest, curiosity, and learning in physics education, J. Sci. Educ. Technol. **27**, 385 (2018).

[20] J. Bennett, F. Lubben, and S. Hogarth, Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching, Sci. Educ. **91**, 347 (2007).

[21] P. J. Fensham, Real world contexts in PISA science: Implications for context-based science education, J. Res. Sci. Teach. **46**, 884 (2009).

[22] M. S. Ubben, F. E. Kremer, S. Heinicke, A. Marohn, and S. Heusler, Smartphone usage in science education: A systematic literature review, Educ. Sci. **13**, 345 (2023).

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

[24] C. C. Tossell, P. Kortum, C. Shepard, A. Rahmati, and L. Zhong, You can lead a horse to water but you cannot make him learn: Smartphone use in higher education, Br. J. Educ. Technol. **46**, 713 (2014).

[25] L. P. Beland and R. Murphy, III communication: Technology, distraction & student performance, Labour Econ. **41**, 61 (2016).

[26] J. M. Van Bruggen, P. A. Kirschner, and W. Jochems, External representation of argumentation in CSCL and the management of cognitive load, Learn. Instr. **12**, 121 (2002).

[27] P. Klein, J. Kuhn, and A. Müller, Förderung von repräsentationskompetenz und experimentbezug in den vorlesungsbegleitenden übungen zur experimentalphysik—Empirische untersuchung eines videobasierten aufgabenformates, Z. Didakt. Naturwiss. **24**, 1 (2018).

[28] P. Klein, Konzeption und untersuchung videobasierter aufgaben im rahmen vorlesungsbegleitender übungen zur experimentalphysik (mechanik), Ph.D. thesis, Technical University of Kaiserslautern, 2016.

[29] A. Kaps, T. Splith, and F. Stallmach, Implementation of smartphone-based experimental exercises for physics courses at universities, Phys. Educ. **56**, 035004 (2021).

[30] A. Kaps and F. Stallmach, Development and didactic analysis of smartphone-based experimental exercises for the smart physics lab, Phys. Educ. **57**, 045038 (2022).

[31] K. Hochberg, S. Becker, M. Louis, P. Klein, and J. Kuhn, Using smartphones as experimental tools—A follow-up: Cognitive effects by video analysis and reduction of cognitive load by multiple representations, J. Sci. Educ. Technol. **29**, 303 (2020).

[32] J. Kuhn and A. Müller, Context-based science education by newspaper story problems: A study on motivation and learning effects, Perspect. Sci. **2**, 5 (2014).

[33] J. K. Gilbert, A. M. W. Bulte, and A. Pilot, Concept development and transfer in context-based science education, Int. J. Sci. Educ. **33**, 817 (2011).

[34] Organisation for Economic Co-operation and Development (OECD), *PISA 2006. Science Competencies for Tomorrow's World. Volume 1: Analysis* (OECD Publishing, Paris, 2007).

[35] National Research Council, *Next Generation Science Standards: For States, By States* (The National Academies Press, Washington, DC, 2013), 10.17226/18290.

[36] M. Ratcliffe and R. Millar, Teaching for understanding of science in context: Evidence from the pilot trials of the twenty first century science courses, J. Res. Sci. Teach. **46**, 945 (2009).

[37] S. Swarat, A. Ortony, and W. Revelle, Activity matters: Understanding student interest in school science, J. Res. Sci. Teach. **49**, 515 (2012).

[38] S. Hidi and W. Baird, Interestingness—A neglected variable in discourse processing, Cognit. Sci. **10**, 179 (1986).

[39] S. Hidi and K. A. Renninger, The four-phase model of interest development, Educ. Psychol. **41**, 111 (2006).

[40] A. Krapp, Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective, Learn. Instr. **12**, 383 (2002).

[41] M. P. Arnone, R. V. Small, S. A. Chauncey, and H. P. McKenna, Curiosity, interest and engagement in technology-pervasive learning environments: A new research agenda, Educ. Technol. Res. Dev. **59**, 181 (2011).

[42] L. M. Gutman and I. Schoon, *The Impact of Non-Cognitive Skills on Outcomes for Young People* (Education Endowment Foundation, London, 2013).

[43] J. C. Valentine, D. L. DuBois, and H. Cooper, The relation between self-beliefs and academic achievement: A meta-analytic review, Educ. Psychol. **39**, 111 (2004).

[44] M. E. Uguroglu and H. J. Walberg, Motivation and achievement: A quantitative synthesis, Am. Educ. Res. J. **16**, 375 (1979).

[45] E. Wild, M. Hofer, and R. Pekrun, Psychologie des lerners, in *Pädagogische Psychologie—Ein Lehrbuch*, edited by A. Krapp and B. Weidenmann (Beltz Psychologie Verlags Union, Weinheim, 2001), pp. 207–270.

[46] R. E. Mayer and R. Moreno, Nine ways to reduce cognitive load in multimedia learning, Educ. Psychol. **38**, 43 (2003).

[47] R. Moreno, Instructional technology: Promise and pitfalls, in *Technology-Based Education: Bringing Researchers and Practitioners Together*, edited by L. PytlikZillig, M. Bodvarsson, and R. Bruning (Information Age Publishing, Greenwich, 2005), pp. 1–19.

[48] R. Moreno and R. E. Mayer, Interactive multimodal learning environments, Educ. Psychol. Rev. **19**, 309 (2007).

[49] R. E. Mayer, Multimedia aids to problem-solving transfer, Int. J. Educ. Res. **31**, 611 (1999).

[50] R. E. Mayer, The promise of multimedia learning: Using the same instructional design methods across different media, Learn. Instr. **13**, 125 (2003).

[51] R. E. Mayer, Cognitive theory of multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, New York, 2005).

[52] J. Sweller, J. J. G. van Merriënboer, and F. G. Paas, Cognitive architecture and instructional design, Educ. Psychol. Rev. **10**, 251 (1998).

[53] J. M. Clark and A. Paivio, Dual coding theory and education, Educ. Psychol. Rev. **3**, 149 (1991).

[54] R. E. Mayer, *Multimedia Learning* (Cambridge University Press, Cambridge, 2009).

[55] S. Alsop and M. Watts, Science education and affect, Int. J. Sci. Educ. **25**, 1043 (2003).

[56] *Multiple Representations in Physics Education.*, edited by D. Treagust, R. Duit, and H. E. Fischer (Springer, Dordrecht, Netherlands, 2017).

[57] J. Scheid, A. Müller, R. Hettmannsperger, and W. Schnotz, Improving learners' representational coherence ability with experiment-related representational activity tasks, Phys. Rev. Phys. Educ. Res. **15**, 010142 (2019).

[58] J. J. G. Van Merriënboer and J. Sweller, Cognitive load theory and complex learning: Recent development and future directions, Educ. Psychol. Rev. **17**, 147 (2005).

[59] R. B. Kozma and J. Russell, Students becoming chemists: Developing representational competence, in *Visualization in Science Education*, edited by J. Gilbert (Springer, Dordrecht, Netherlands, 2005).

[60] 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).

[61] E. Torigoe and G. Gladding, Connecting symbolic difficulties with failure in physics, Am. J. Phys. **79**, 133 (2011).

[62] T. Seufert, Supporting coherence formation in learning from multiple representations, Learn. Instr. **13**, 227 (2003).

[63] R. E. Mayer, Systematic thinking fostered by illustrations in scientific text, J. Educ. Psychol. **81**, 240 (1989).

[64] R. E. Mayer, Rote versus meaningful learning, Theory Pract. **41**, 226 (2002).

[65] C. B. Fried, In-class laptop use and its effects on student learning, Comput. Educ. **50**, 906 (2008).

[66] A. Kaps, T. Splith, and F. Stallmach, Implementation of smartphone-based experimental exercises for physics courses at universities, Phys. Educ. **56**, 035004 (2021).

[67] A. Kaps and F. Stallmach, Development and didactic analysis of smartphone-based experimental exercises for the smart physics lab, Phys. Educ. **57**, 045038 (2022).

[68] M. McCloskey, Naïve theories of motion, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Erlbaum, Hillsdale, NJ, 1983), pp. 299–324.

[69] J. J. Clement, Students' preconceptions in introductory mechanics, Am. J. Phys. **50**, 66 (1982).

[70] L. C. McDermott, Research on conceptual understanding in mechanics, Phys. Today **37**, No. 7, 24 (1984).

[71] *Physics Education*, edited by H. E. Fischer and R. Girwidz (Springer, Cham, 2021).

[72] A. Madsen, S. B. McKagan, and E. C. Sayre, MMS17 resource letter RBAI-1: Research-based assessment instruments in physics and astronomy, Am. J. Phys. **85**, 245 (2017).

[73] A. Gasparini, Using mobile devices as experimental tools in physics lessons: An empirical study of the effects on learning and motivation at secondary school level, Ph.D. thesis, University of Geneva, 2021, https://archive-ouverte.unige.ch/unige:155124.

[74] D. Hestenes, M. Wells, and W. Swackhamer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992).

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

[76] M. T. Svec, Effect of micro-computer based laboratory on graphing interpretation skills and understanding of motion, in *presented at the Annual Meeting of the National Association for Research in Science Teaching, San Francisco, CA* (1995).

[77] A. Helmke and F.-W. Schrader, School achievement, cognitive and motivational determinants, in *International Encyclopedia of the Social and Behavioural Sciences* (Elsevier, Oxford 2001), Vol. 20, pp. 13552–13556.

[78] C. Bolte and H. Stork, Entwicklung und einsatz von erhebungsinstrumenten zur analyse der schüler-lehrer- interaktion im chemieunterricht (teil I), Z. Didakt. Naturwiss. **2**, 21 (1996).

[79] C. Bolte, Entwicklung und einsatz von erhebungsinstrumenten zur analyse der schüler-lehrer-interaktion im chemieunterricht (teil II), Z. Didakt. Naturwiss. **3**, 67 (1997).

[80] F. D. Naylor, A State-Traint Curiosity Inventory. Effect of computer graphics on improving estimates to algebra word problems, J. Educ. Psychol. **77**, 285 (1981).

[81] J. A. Litman, R. P. Collins, and C. D. Spielberger, The nature and measurement of sensory curiosity, Pers. individ. Differ. **39**, 1123 (2005).

[82] R. B. Elkstrom, J. W. French, and H. H. Harman, *Manual for Kit of Factor-Referenced Cognitive Tests* (Educational Testing Service, Princeton, 1976), p. 176.

[83] M. S. Khine, *Visual-Spatial Ability in STEM Education* (Springer International Publishing, Cham, 2017).

[84] M. Opfermann, A. Schmeck, and H. E. Fischer, Multiple representations in physics and science education—Why should we use them?, in *Multiple Representations in Physics Education*, edited by D. Treagust, R. Duit, and H. E. Fischer (Springer, Dordrecht, Netherlands, 2017).

[85] H. W. Marsh, The structure of academic self-concept: The Marsh/Shavelson model, J. Educ. Psychol. **82**, 623 (1990).

[86] T. Ritcher, J. Naumann, and N. Groeben, Attitudes toward the computer: Construct validation on an instrument with scales differentiated by content, Comput. Hum. Behav. **16**, 473 (2000).

[87] M. Luckay and B. I. Collier-Reed, Admitting engineering students with the best chance of success: Technology literacy and the Technological Profile Inventory (TPI), in *Proceedings of the 1st Biennial Conference of the South African Society for Engineering Education* (Stellenbosch, 2011).

[88] N. Orion, A. Hofstein, P. Tamir, and G. J. Giddings, Development and validation of an instrument for assessing the learning environment of outdoor science activities, Sci. Educ. **81**, 161 (1997).

[89] M. L. Yunker, A systemic examination of the introduction of an outdoor learning-based science curriculum to students, their teacher, and the school principal, Ph.D. thesis, University of Michigan, 2010.

[90] F. G. Paas, J. J. Van Merrienboer, and J. J. Adam, Measurement of cognitive load in intructional research, Percept. Mot. Skills **79**, 419 (1994).

[91] R. Cors, Informal science learning: An investigation of how novelty and motivation affect interest development at a mobile laboratory, Ph.D. thesis, University of Geneva, 2016.

[92] M. Hirth, Akustische untersuchungen mit dem smartphone und tablet-computern-fachliche und didaktische aspekte, Ph.D. thesis, Technical University of Kaiserslautern, 2019.

[93] J. Woithe, A. Müller, S. Schmeling, and J. Kuhn, Motivational outcomes of the science outreach lab S'Cool LAB at CERN: A multilevel analysis, J. Res. Sci. Teach. **59**, 930 (2022).

[94] M. Hänze and R. Berger, Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes, Learn. Instr. **17**, 29 (2007).

[95] J. P. Dorman, Cross-national validation of the what is happening in this class? (WIHIC) questionnaire using confirmatory factor analysis, Learn. Environ. Res. **6**, 231 (2003).

[96] B. J. Fraser, C. J. McRobbie, and D. Fischer, Development, validation and use of personal and class forms of a new classroom environment questionnaire, in *Proceedings of the Western Australian Institute for Educational Research Forum* (1996).

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

[98] P. Kline, *A Handbook of Test Construction: Introduction to Psychometric Design* (Routledge, London, 2015).

[99] J. M. Cortina, What is coefficient alpha? An examination of theory and applications, J. Appl. Psych. **78**, 98 (1993).

[100] P. Cohen, J. Cohen, L. S. Aiken, and S. G. West, The problem of units and the circumstance for POMP, Multivar. Behav. Res. **34**, 315 (1999).

[101] B. G. Tabachnick and L. S. Fidell, *Experimental Designs Using ANOVA* (Thomson/Brooks/Cole, Belmont, CA, 2007).

[102] L. Ding and X. Liu, Getting started with quantitative methods in physics education research, in *Getting Started in PER*, Reviews in PER Vol. 2, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2012), 10.1119/RevPERv2.3.2.

[103] A. Field, J. Miles, and Z. Field, *Discovery Statistics Using R* (SAGE publications, Los Angeles, 2012).

[104] *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*, edited by P. Velentgas, N. A. Dreyer, P. Nourjah, S. R. Smith, and M. M. Torchia (Agency for Healthcare Research and Quality, Rockville, 2013).

[105] Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B **57**, 289 (1995).

[106] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge Academic, New York, 1988).

[107] A. C. Hattie, *Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (Routledge, London, New York, 2009).

[108] R Core Team, R: A language and environment for statistical computing (version 4.2.2) [Computer software], Retrieved from https://cran.r-project.org (2022).

[109] W. K. Adams and C. E. Wieman, Development and validation of instruments to measure learning of expert-like thinking, Int. J. Sci. Educ. **33**, 1289 (2011).

[110] K. S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. **48**, 1273 (2018).
[111] M. R. Semak, R. D. Dietz, R. H. Pearson, and C. W. Willis, Examining evolving performance on the Force Concept Inventory using factor analysis, Phys. Rev. Phys. Educ. Res. **13**, 010103 (2017).
[112] M. L. van Blerkom, *Measurement and Statistics for Teachers* (Routledge, London, 2009).
[113] M. D. Miller, R. L. Linn, and N. E. Gronlund, *Measurement and Assessment in Teaching* (Pearson, Upper Sadler River, 2009).
[114] A. Evers, C. Hagemeister, and A. Hostmaelingen, *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests* (European Federation of Psychology Associations, Brussels, 2013).
[115] T. J. Kline, Classical test theory: Assumptions, equations, limitations, and item analyses, in *Psychological Testing: A Practical Approach to Design and Evaluation* (SAGE Publications, Thousand Oaks, CA, 2005), pp. 91–106.
[116] V. E. Lee and D. T. Burkam, Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis, Sci. Educ. **80**, 613 (1996).
[117] R. A. Louis and J. M. Mistele, The differences in scores and self-efficacy by student gender in mathematics and science, Int. J. Sci. Math. Educ. **10**, 1163 (2012).
[118] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).
[119] E. M. Smith, M. M. Stein, C. Walsh, and N. G. Holmes, Direct measurement of the impact of teaching experimentation in physics labs, Phys. Rev. X **10**, 011029 (2020).
[120] E. Smith and N. Holmes, Best practice for instructional labs, Nat. Phys. **17**, 662 (2021).
[121] N. G. Holmes and E. M. Smith, Operationalizing the AAPT learning goals for the lab, Phys. Teach. **57**, 296 (2019).
[122] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, Science **332**, 862 (2011).
[123] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).
[124] S. I. Hofer, R. Schumacher, H. Rubin, and E. Stern, Enhancing physics learning with cognitively activating instruction: A quasi-experimental classroom intervention study, J. Educ. Psychol. **110**, 1175 (2018).
[125] L. Steidtmann, T. Kleickmann, and M. Steffensky, Declining interest in science in lower secondary school classes: Quasi-experimental and longitudinal evidence on the role of teaching and teaching quality, J. Res. Sci. Teach. **60**, 164 (2023).
[126] R. Karam, Thematic issue: The interplay of physics and mathematics: Historical, philosophical and pedagogical considerations, Sci. Educ. **24**, 487 (2015).
[127] *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B. Eylon (Springer, New York, 2019).
[128] A. Thorndike, Correlation between physics and mathematics grades, Sch. Sci. Math. **46**, 650 (1946).
[129] D. E. Meltzer, The relationship between mathematics preparation, and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores, Am. J. Phys. **70**, 1259 (2002).
[130] L. Cui, N. S. Rebello, and A. G. Bennett, College students' transfer from calculus to physics, AIP Conf. Proc. **818**, 37 (2006).
[131] 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).
[132] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19**, 479 (2001).
[133] 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).
[134] A. Siani, BYOD strategies in higher education: Current knowledge, students' perspectives, and challenges, New Dir. Teach. Phys. Sci. **12**, 1 (2017).
[135] T. Cochrane, L. Antonczak, H. Keegan, and V. Narayan, Riding the wave of BYOD: Developing a framework for creative pedagogies, Res. Learn. Tech. **22**, 24637 (2014).
[136] H. Lin, Implementing large-scale mobile device initiatives in schools and institutions, in *Emerging Technologies for STEAM Education: Full STEAM Ahead* (Springer, 2015), pp. 179–198.
[137] P. Onorato, L. M. Gratton, M. Polesello, A. Salmoiraghi, and S. Oss, The Beer Lambert law measurement made easy, Phys. Educ. **53**, 035033 (2018).
[138] *Handbook of Mobile Teaching, and Learning*, edited by A. Zhang and D. Cristol (Springer, Berlin, 2015), Chap. 38.
[139] A. Bouchereau, Les objets connectés au service de l'apprentissage, Ph.D. thesis, Université Bourgogne Franche-Comté, 2020.
[140] T. De Jong, Cognitive load theory, educational research, and instructional design: Some food for thought, Instr. Sci. **38**, 105 (2010).
[141] W. Schnotz and C. Kürschner, A reconsideration of cognitive load theory, Educ. Psychol. Rev. **19**, 469 (2007).
[142] K. E. DeLeeuw and R. E. Mayer, A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load, J. Educ. Psychol. **100**, 223 (2008).
[143] V. Kenttälä, From design to use—factors of value creation in the usability and implementation of digital learning technology, Ph.D. thesis, University of Jyväskylä: Faculty of Information Technology, 2019.
[144] P. Bihouix and K. Mauvilly, *Le Désastre de L'école Numérique* (Ed. Seuil, Paris, 2021).
[145] M. Kratz, *Experimente als Hausaufgaben—Physik* (Aulis-Verlag Deubner, Köln, 2004).
[146] Institute of Physics, Physics education at home: DIY activities for home learning, Website, https://iopscience

.iop.org/journal/0031-9120/page/physics-education-at-home (2020).

[147] P. Klein, L. Ivanjek, M. N. Dahlkemper, K. Jelicic, M.-A. Geyer, S. Küchemann, and A. Sušac, Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories, Phys. Rev. Phys. Educ. Res. **17**, 010117 (2021).

[148] K. Jelicic, M.-A. Geyer, L. Ivanjek, P. Klein, S. Küchemann, M. N. Dahlkemper, and A. Sušac, Lab courses for prospective physics teachers: What could we learn from the first COVID-19 lockdown?, Eur. J. Phys. **43**, 055701 (2022).

[149] S. Z. Lahme, P. Pirinen, L. Rončevič, A. Lehtinen, A. Sušac, P. Klein, and A. Müller, Physics lab courses under digital transformation: A survey among German, Finnish, and Croatian university lab instructors about the role of new digital technologies and learning objectives, arXiv: 2305.08515.

[150] G. Pospiech, Framework of mathematization in physics from a teaching perspective, in *Mathematics in Physics Education*, edited by G. Pospiech, M. Michelini, and B. Eylon (Springer, Cham, Switzerland, 2019).

[151] S. J. Wolf and B. J. Fraser, Learning environment, attitudes and achievement among middle-school science students using inquiry-based laboratory activities, Res. Sci. Educ. **38**, 321 (2008).

[152] B. J. Fraser, Classroom learning environments: Historical, and contemporary perspectives, in *Handbook of Research on Science Education*, edited by N. G. Ledermann, and S. K. Abel (Routledge, 2014), Vol. II, pp. 118–133.

[153] C. Angeli and N. Valanides, *Technological Pedagogical Content Knowledge: Exploring Developing and Assessing TPCK* (Springer, New York, 2015).

[154] D. Fanelli, Negative results are disappearing from most disciplines and countries, Scientometrics **90**, 891 (2012).

[155] A. Franco, N. Malhotra, and G. Simonovits, Publication bias in the social sciences: Unlocking the file drawer, Science **345**, 1502 (2014).

[156] L. D. Conlin, E. Kuo, and N. R. Hallinen, How null results can be significant for physics education research, Phys. Rev. Phys. Educ. Res. **15**, 020104 (2019).

[157] J. W. Osborne, A. B. Costello, and J. T. Kellow, Best practices in exploratory factor analysis, in *Best Practices in Quantitative Methods*, edited by J. Osborne (SAGE Publications, Los Angeles, 2012), p. 86.

[158] K. Schermelleh-Engel, H. Moosbrugger, and H. Müller, Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures, Methods Psychol. Res. Online **8**, 23 (2003).

[159] B. A. Cerny and H. F. Kaiser, A study of a measure of sampling adequacy for factor-analytic correlation matrices, Multivariate Behav. Res. **12**, 43 (1977).

[160] J. P. Stevens, *Applied Multivariate Statistics for the Social Sciences* (Lawrence Erlbaum, Mahwah, NJ, 2002).

[161] S. K. Reed, Effect of computer graphics on improving estimates to algebra word problems, J. Educ. Psychol. **77**, 285 (1985).

[162] S. K. Reed and N. C. Saavedra, A comparison of computation, discovery, and graph procedures for improving students' conception of average speed, Cognit. Instr. **3**, 31 (1986).