Journal of Science Education and Technology (2018) 27:385–403
https://doi.org/10.1007/s10956-018-9731-7

Using Smartphones as Experimental Tools—Effects on Interest, Curiosity, and Learning in Physics Education

Katrin Hochberg[1] · Jochen Kuhn[1] · Andreas Müller[2]

Abstract

Smartphones as experimental tools (SETs) offer inspiring possibilities for science education, as their built-in sensors allow many different measurements, but until now, there has been little research that studies this approach. Due to current interest in their development, it seems necessary to provide empirical evidence about potential effects of SETs by a well-controlled study. For the present investigation, experiments were developed that use the smartphones' acceleration sensors to investigate an important topic of classical mechanics (pendulum). A quasi-experimental repeated-measurement design, consisting of an experimental group using SETs (smartphone group, SG, $N_{SG} = 87$) and a control group working with traditional experimental tools (CG, $N_{CG} = 67$), was used to study the effects on interest, curiosity, and learning achievement. Moreover, various control variables were taken into account. With multiple-regression analyses and ANCOVA, we found significantly higher levels of interest in the SG (small to medium effect size). Pupils that were less interested at the beginning of the study profited most from implementing SETs. Moreover, the SG showed higher levels of topic-specific curiosity (small effect size). No differences were found for learning achievement. This means that the often-supposed cognitive disadvantage of distracting learners with technological devices did not lead to reduced learning, whereas interest and curiosity were apparently fostered. Moreover, the study contributes evidence that could reduce potential concerns related to classroom use of smartphones and similar devices (increased cognitive load, mere novelty effect). In sum, the study presents encouraging results for the under-researched topic of SET use in science classrooms.

Keywords Smartphones · Technology-based activities · Secondary education · Physics education · Interest · Curiosity

Introduction

Smartphones are nowadays ubiquitous: 73% of teens (aged 13 to 17) in the United States report to have or at least have access to smartphones (Lenhart 2015). As these devices come with a multitude of built-in sensors, they cannot only be used for communication or browsing the internet, but also as mobile pocket-labs. With the internal sensors, numerous variables, such as acceleration, sound pressure level, frequency, illuminance, magnetic flux density, or even ionizing radiation, can be measured anywhere and anytime, creating new possibilities for teaching and learning, especially in science education. Apart from simple research or documentation, the built-in sensors of smartphones allow one to use the devices as experimental tools. The advantages lie not in the ability to take more accurate measurements, but in the ability to do a variety of experiments and interpret those quickly and easily. With the multitude of internal sensors, it is possible to cover phenomena in mechanics, acoustics, electromagnetism, optics, and radioactivity (by using the CMOS sensor of the camera; Kuhn et al. 2014) with only one device. Applications (apps) generate tables, graphs, or other forms of data representation automatically on the smartphone screen. Furthermore, the portability and versatility of SETs allows for experiments inside as well as outside the classroom and also for experiments as homework (as almost

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10956-018-9731-7) contains supplementary material, which is available to authorized users.

✉ Katrin Hochberg
khochberg@physik.uni–kl.de

[1] Department of Physics/Physics Education Research Group, University of Kaiserslautern, Erwin-Schrödinger-Straße 46, D-67663 Kaiserslautern, Germany

[2] Faculty of Sciences/Physics Section and Institute of Teacher Education, Pavillon d'Uni Mail (IUFE), University of Geneva, Boulevard du Pont d'Arve 40, 1211 Geneva, Switzerland

every pupil possesses their own smartphone). Thus, experiment-oriented seamless learning can be generated. Numerous *experimental* concepts for smartphone use in science education have appeared in the last years (kinematics: Koleza and Pappas 2008; oscillations and waves: Castro-Palacio and Velázquez-Abad 2013; Parolin and Pezzi 2013; Sans et al. 2013; other areas of mechanics: Chevrier et al. 2013; Hochberg et al. 2014; Kuhn and Vogt 2013; Monteiro et al. 2014; Shakur and Sinatra 2013; acoustics: Greenslade 2016; Hirth et al. 2015; Müller et al. 2016; Vogt et al. 2015; optics: Klein et al. 2014; Thoms et al. 2013; electromagnetism: Forinash and Wisman 2012; Silva 2012; and radioactivity: Kuhn et al. 2014). The Physics Teacher journal has established a column about SETs since 2012 (Kuhn and Vogt 2012).

The *educational* effects of using smartphones in school, however, remain a controversial issue in current discussions (see, e.g., Barkham and Moss 2012). Although recent studies indicate that technology can be an effective way to enhance pupils interest, further study is needed in order to seamlessly integrate technology with factors such as instructional materials and teaching contexts (Swarat et al. 2012). Moreover, it is a well-known concern that new technology in educational settings may lead to distraction or increased cognitive load and thus impair learning (van Bruggen et al. 2002; Fried 2008; Tossell et al. 2014).

Lack of empirical evidence in general, as well as possible drawbacks and aversive effects in particular can lead to a serious impediment of the development of a given instructional technology, or as Newhouse and Rennie (2001) warn, that even when there are compelling arguments in favor of some approach, “the field of educational technology is littered with discarded technologies which had equally compelling support.” Unfortunately, in spite of the extensive literature about the experimental possibilities, there have been few empirical studies regarding the learning effects of SETs in science education to support this discussion (see, e.g., Kuhn and Vogt 2015; Mazzella and Testa 2016). Despite the seeming “boom” of SETs due to favorable theoretical arguments, timely empirical evidence is necessary. In this contribution, such an empirical investigation of smartphones as experimental tools is presented.

Theoretical Background and Rationale

Theoretical Framework

Fostering Interest with Authentic Devices

In the present study, we investigated SETs as a potential special kind of contextualized science learning, in line with the theoretical framework of context-based science education (CBSE). In order to improve learning and motivation, CBSE calls for authentic contexts, which in the sense of a widespread understanding (starting with the word origin: gr. *authentikós* “true“) are related to actual, real(istic), genuine situations, and experiences learners are supposed to encounter. This is, for example, the understanding assumed by PISA 2006 (Organisation for Economic Co-operation and Development [OECD] 2007) which repeatedly states the usage and value of tasks and problems “that could be part of the actual experience or practice of the participant in some real-world setting” (OECD 2007, p. 81), and it “places most value on tasks that could be encountered in a variety of real-world situations” (OECD 2007, p. 81). Such contexts can open up an enclosed synthetic reality to be found in many physics classes, in which pupils learn things that seem to have no relation to the real world and their everyday life, resulting in questioning the reasons for studying the subject at all (Müller 2006; Taasoobshirazi and Carr 2008). Existing empirical studies show that a provision of relevance by an authentic context can lead to increased motivation and interest: A systematic review (Bennett et al. 2007; see also Gilbert et al. 2011) found evidence that certain context-based approaches promote understanding as well as conventional approaches, while fostering more positive attitudes to science. The reviewed studies were using approaches where contexts were used as starting point for the development of scientific ideas or where links between science, technology, and society were emphasized (science-technology-society or STS approaches). Fifteen of the 17 reviewed studies investigated interventions that lasted for at least 1 year. As Ratcliffe and Millar (2009) argued, a difficulty of such fully context-based science curricula is that many teachers are challenged by the need to take on far reaching changes of pedagogic strategies. Thus, to enhance feasibility and flexibility, interventions on a smaller scale of curricular and pedagogical change have been proposed and investigated, at the level of instructional episodes, in the sense of a “segment of instruction devoted to a specific content topic or skill” (Swarat et al. 2012, p. 520), and representing “an integral and independent potential learning event (…) (e.g., about the motion of a mechanical pendulum) resulting from and constrained by the content and structure of the instructional materials, as well as by the intentions of a teacher” (Swarat et al. 2012, p. 520). These interventions do not require special methodological or organizational conditions and can easily be embedded in existing teaching approaches. Examples for context-based interventions on the instructional episode level are newspaper story problems or other context-based instruction material, which were shown to improve motivation and learning in comparison to traditional classes with medium to large effect sizes ($\omega^2 = 0.52$ for motivation, $\omega^2 = 0.20$ for learning, see Kuhn and Müller 2014, $d = 1.52$ for learning, see Bahtaji 2015).

Inspired by the success of CBSE interventions on the level of instructional episodes, we tried to further reduce the complexity of the context and hence increase the easiness of its embedding in existing teaching approaches, while still fostering motivation and learning. As smartphones are ubiquitous for pupils nowadays, the devices themselves may be able to generate a relevant context in the following way: Experimenting with smartphones means experimenting with a device that is familiar and important to pupils, using an app that they could use on their own device outside of the classroom, too. We call this connection of the experimental medium itself, based on its material basis, to pupils' everyday life a "material" context. In addition to this material context that SETs may provide, they can be used for CBSE in the classical sense. With SETs, mobile experiments can be conducted, investigating authentic topics in everyday life wherever and whenever they appear (see the "Introduction" section and in particular Monteiro et al. 2016; Müller et al. 2016; Müller et al. 2015; Tornaría et al. 2014, for some recent applications to everyday life phenomena). We call this more traditional kind of context "topical context." According to the theoretical framework of CBSE, this topical context can generate positive affective and cognitive effects. The material context might also affect motivation and learning, as the devices themselves represent a link to everyday life, especially for youths, even if SETs are used in otherwise conventional experimental setups. As context in mobile experiments may be generated by the experimental medium itself as well as by the tasks and the topics (material and topical context), there may be a twofold relation of the content to pupils' everyday life. Hence, higher affective and cognitive effects can be expected than the effects that might be created by the material context alone.

Beyond the practical advantages (see the "Introduction" section) and the benefits predicted by the theoretical framework of CBSE, using SETs is also promising according to another strand of research: Swarat et al. (2012) investigated the role of learning environment elements (content topics, activities, and learning goals, in their terms) on pupils' interest in school science. A material context can be considered a specific kind of "activity" in their sense, found to elicit interest by its "hands on" nature and engagement with technology, provided it is effectively integrated with factors such as instructional materials and teaching contexts. As Swarat et al. (2012) say, "technology may enhance student interest by connecting students with real data and thus promoting a sense of authenticity". This is very much in the sense of the specific aspect of CBSE we address by "material context," the effect of "authentic data" further supported by the "authenticity" of smartphone devices for young people. Moreover, a topical context in the sense introduced above is closely related to "content topic" in terms of Swarat et al. (2012).

In the current study, interest as a key component of motivation was investigated, Hidi and Renninger (2006) define interest as "psychological state of engaging or the predisposition to reengage with particular classes of objects, events, or ideas over time" (Hidi and Renninger 2006, p. 112). We focus on interest, because "in contrast to many other motivational concepts, interest is characterized by its content or object specificity" (Krapp 2005, p. 382). For a meaningful use of SETs in physics teaching, it is important that interest in the physics behind the experiments is increased ("content specificity"). If pupils are only interested in the devices themselves ("object specificity"), this would probably lead to motivation fading soon after the activity is done. In accord with the four-phase model of interest development by Hidi and Renninger (2006), using SETs can trigger situational interest, because pupils conduct a "hands on," technology-enhanced activity in the sense of Swarat et al. (2012). This interest can evolve into a maintained situational interest, when it is "held and sustained through meaningfulness of tasks and/or personal involvement" (Harackiewicz et al. 2000; Mitchell 1993) (Hidi and Renninger 2006, p. 114). Material contexts as provided by SETs may be meaningful to pupils because they know the devices from their everyday life and have a strong connection with them; hence, SETs can foster object-specific-maintained situational interest. In order to turn this object-specific interest into a content-specific, emerging individual interest for physics, pupils need a context to acknowledge the relevance of the studied topics in accord with the theoretical framework of CBSE. We believe that the mere material context of smartphones used in otherwise traditional experiments may already lead to this acknowledgement (see preceding texts), to an increased interest in the topics and hence to an emerging individual interest for physics, which is not object specific anymore, but content specific. In short, SETs as a "hands on" activity can increase a situational, object-specific interest, while at the same time, by providing a material context they might transfer this interest into a content-specific, individual interest.

Increasing Curiosity by Teaching Pupils to Collect Data with SETs

In addition to interest, we studied the "motivational internal state" (Berlyne 1978, p. 144) of curiosity. Recent research (von Stumm et al. 2011) showed that curiosity, correlated with effort, is an important predictor of academic achievement and hence has been considered in an increasing number of studies in the last few years. According to Arnone et al. (2011), curiosity is "a desire for new information or experience" (Arnone et al. 2011, p. 185), which includes a trigger, a reaction to that trigger, and a resolution (of the desire), which can be satisfactory or unsatisfactory. A satisfactory resolution can initiate new learning, interest, and engagement, which can re-trigger curiosity. Thereby, learners are only motivated to seek information and persist in exploration until they reach a satisfactory resolution, if they value the prospect of satisfying curiosity and believe to have the competence to reach this satisfactory resolution. The perceptions of value and expectancy for success are moderated by personal and contextual or situational factors. Personal factors are, for example, curiosity as a personality trait, as well as the self-concept of the learner and cognitive abilities. Contextual factors refer to the learning setting, e.g., formal or informal. Situational factors include "all those factors 'in the moment' that help explain the direction of behaviour" (Arnone et al. 2011, p. 190), in particular "object" and "content specificities" as explained above. If learners believe that "information seeking cannot be satisfied in a timely manner, with minimal effort" (Arnone et al. 2011, p. 191), a satisfactory resolution of curiosity and hence increase in interest may not occur. By introducing SETs, we teach pupils how to use the familiar devices in a new and unexpected way as mobile pocket-labs. Pupils are provided with a new way of accessing information in those quite effortlessly produced data. Thus, pupils' perceived competence in quickly, easily and successfully conducting experiments with their own smartphones is increased. This should, in turn, raise their confidence in being able to find a satisfactory resolution, thereby increasing their curiosity.

Learning with SETs: Beneficial or Distracting?

A further hypothesis behind approaches to support learners motivation (such as CBSE) is the widespread belief in education that better motivation will lead to better learning (see, e.g., Bennett et al. 2007). However, correlations between motivation and learning are around $r \approx 0.3$ (Uguroglu and Walberg 1979; Wild et al. 2001), i.e., lower than one might expect. Nevertheless, an increase in learning achievement by using SETs is predicted by theoretical frameworks different from CBSE, in particular cognitive load theory (Sweller et al. 1998), which takes into account the fact that the working memory has limited capacities. Three different forms of cognitive load occur in learning processes: intrinsic cognitive load, generated by the nature of the learning content, extraneous cognitive load, generated by the presentation of the learning material, and germane cognitive load, generated by meaningful learning. The sum of all three kinds of cognitive load cannot exceed the capacities of working memory. Hence, instruction design should strive to reduce the extraneous cognitive load, doing so should leave more free cognitive capacity for the germane cognitive load and thereby increase learning achievement. As in most studies, extraneous cognitive load is manipulated and is commonly referred to as "the" cognitive load in the literature and also in this paper, if not specified differently. SETs can reduce extraneous cognitive load, e.g., by providing graphs automatically: Pupils do not have to expend cognitive capacities by drawing graphs themselves (extraneous load), but can concentrate fully on their interpretation (germane load). This is also in agreement with the strong current of research and development concerning technology-enhanced learning, where since the early 1990s (see, e.g., Thornton and Sokoloff 1990), we saw new approaches to science education, where routine jobs got increasingly computerized. As Tho et al. (2015) put it, this "allows more time for the students to carry out other educationally meaningful tasks and activities such as analysis and interpretation of results, redoing the experiments for testing alternative hypotheses". Additionally, cognitive capacities are used more effectively with higher motivation, which is particularly sparked by new media (Cognitive–Affective Theory of Learning with Media, Moreno 2005). Moreover, graphs, tables, and additional kinds of representations (dependent of the app and the sensor used) are produced easily with SETs and can be connected with each other and the observation of the experiment, which, according to accounts of effective use of multiple representations, can also foster learning achievement (Ainsworth 2006). Real-time graphs can also be used to challenge intuitive conceptions and support a better understanding of kinematics (Beichner 1996; Brasell 1987). Another framework that provides reasons for a beneficial use of SETs is that of mobile learning. Mobile learning, defined as "the processes of coming to know through conversations across multiple contexts among people and personal interactive technologies" (Sharples et al. 2007, p. 224) provides the possibility of personalized and contextualized learning, less restricted by temporal or environmental constraints than any other learning technology before (Crompton et al. 2016). More specifically, a number of theoretical and empirical arguments are put forward in favor of mobile learning. On the affective level, learner's attitudes and perceptions towards by learning can be improved by mobile and technology enhanced approaches (Wu et al. 2012; Swarat et al. 2012). On the cognitive level, mobile technology can support and enrich learning in various ways (Haßler et al. 2016): (a) flexible and adaptive learning (learners respond and react to the learning environment, which in turn is no longer predetermined); (b) meaningful and situated learning, and (c) personalized learning. Recent reviews and meta-analyses discuss these educational opportunities in terms of critical success factors (Alrasheedi et al. 2015), general learning theories (Crompton et al. 2017) and other characteristics (Hwang and Tsai 2011; Wu et al. 2012; Pimmer et al. 2016). An interesting finding in the present context is that, within science education, a majority of studies was carried out in informal settings in the life sciences (Crompton et al. 2016), i.e., in an area where a kind of "mobile" learning has already a long tradition ("field trips," "outdoor learning;" Falk and Balling 1979; DeWitt and Storksdieck 2008). Put differently, in physical science, there is yet very little research to explore the educational effects of mobile learning, and the present work is a contribution in that sense.

Yet, in opposition to these theoretical reasons for advantageous effects of using SETs on learning achievement, there are also studies that showed negative effects of using instructional technology in school: Van Bruggen et al. (2002) studied the effect of computer-supported learning environments on cognitive load. They concluded that as pupils have to integrate and coordinate multiple representations during learning (the same applies to using SETs), their attention is split between several sources of information and cognitive load is increased. As elaborated above, this will reduce learning achievement. Another common critique of using smartphones in school is their potentially distracting effect. Tossell et al. (2014) provided smartphones for pupils who had never used such devices before. At the beginning of the study, most participants were confident that the smartphone would help them reach their educational goals. Nevertheless, after 1 year, although pupils had indeed used their smartphones for informal learning, they perceived them as distracting rather than helpful. Another study (Beland and Murphy 2015) investigated pupils' achievement before and after smartphones were banned from their school completely. Performance in high stakes exams significantly increased post ban, especially for low-achieving pupils, which the authors attribute to less distractions by smartphones in class. Potential distracting effects are not only studied regarding the use of smartphones in school, but regarding ICT in general (see for example Fried 2008, regarding laptops). In view of this contradictory state of research, we formulate below an open research question (rather than a directed research question) on potential learning effects of SETs.

Focus on Material Context in the Current Study

Based on the theoretical framework elaborated above, SETs have the potential to foster interest, curiosity, and learning achievement in various ways. We decided to focus the current study on the effects of potential material contexts by introducing SETs in otherwise conventional classroom settings (no authentic topical context), i.e., not to exploit the full potential of combining both kinds of context at the same time. For this, there are two main reasons: Firstly, affective and cognitive effects of using SETs in the classroom have been hardly studied up to now, despite of a booming literature about experimental possibilities (see the "Introduction" section). As a consequence, it is important to manipulate only one group of variables that may influence affective and cognitive effects, while keeping all the others constant, respectively controlled. Secondly, the focus on the material context rather than the topical context is of educational nature: We believe that a formal introduction in the use of SETs in an otherwise conventional classroom setting without any authentic topical context is necessary in order for pupils to be able to use them effectively. Pupils need to learn about the functions of the device in conventional settings to prevent misunderstandings or even misconceptions. For example, due to the operating mode of the acceleration sensor, the smartphone measures an acceleration of g (9.81 m/s^2) while it is at rest. This and other sensor specifications have to be explained to and discussed with pupils in a formal setting before experimenting on their own—especially in mechanics, as this topic is well-known for widespread misconceptions. In the current study, SETs were thus used in a traditional setting with well-known classroom experiments instead of for mobile experiments outside of the classroom. Because the only difference between the interventions in smartphone and control group was the experimental tool pupils were working with and the short duration of the intervention (3 h), affective and cognitive effects are supposed to be rather small. A follow-up test was conducted to check whether the effects would be stable over time, if they would be larger than expected.

Control Variables

In the framework of this investigation, the *perceived relation to reality* of physics classes was assessed: If, as intended, using SETs provides a material context, pupils working with those devices should recognize the relationship of physics to their everyday life more than pupils working with traditional experimental tools.

To ensure that the focus of the study lies in the potential material context provided by SETs and confounding effects were avoided, several variables were kept constant or controlled. First, it is known from several studies, that the teacher has an effect on learning (see, e.g., Hattie 2008). All pupils were therefore taught by the same *tutor*, who was, in turn, evaluated by the pupils to ensure that there was no bias. Second, we tried to keep the *cognitive load*, generated by the written instruction materials, equal for both groups as explained in "The Intervention" section. Thus, any differences in the measured cognitive load were due to the use of SETs as opposed to the traditional experimental tools (see the "Learning with SETs: Beneficial or Distracting?" section for potential positive and negative cognitive effects of using SETs according to existing research). Third, several variables which could not be kept constant over all participants were measured and considered in the analyses: As "interest in physics instruction is closely related to the students' physics-related self-concept" (Hoffmann 2002, p. 452), pupils with high *self-concept* should be more interested in physics. Furthermore, according to the theoretical framework of curiosity detailed above, self-concept and *curiosity as a personality trait* can moderate the expectancy for success in answering questions satisfactorily and hence influence curiosity and interest. In addition to the influence on perceived novelty and the moderating effect on curiosity, pupils' self-concept has a reciprocal relationship with achievement (see, e.g., Hattie 2008; Main and O'Rourke 2011). If by accident, there are more pupils with a high physics-related self-concept in one of the treatment groups, this might lead to unintended group differences, which are not generated by the material context. Hence, it is important to measure this variable and include it in the analyses. This is also true for other variables: As the intervention consists of experiments, for pupils working with, as well as those working without, SETs, more *experience with experiments* will hypothetically lead to a higher experimental competence, which can have a beneficial effect on learning achievement. If there are only little difficulties in conducting the experiments, this effect will decrease, as even pupils without prior experience can succeed in the intervention. According to cognitive load theory, *spatial abilities* can influence the processing of information in working memory (Mayer and Moreno 2003), and so, pupils with high spatial abilities might have advantages regarding learning. Another important variable to include in the analyses is *gender*: Due to their socialization, girls might have less prior experience in tinkering with technological devices and conducting experiments (Osborne et al. 2003), which would lead to a higher novelty of the intervention and hence influence affective and cognitive effects. Studies also found that for various reasons, female pupils have less positive attitudes towards science by the end of the second year of secondary school (Reid and Skryabina 2002) and women show lower achievement in concept tests in physics (see, e.g., Hazel et al. 2007; Kost-Smith et al. 2010). Gender may hence have indirect as well as direct influences on affective and cognitive effects.

Consistent with the above considerations (and described in detail in the section about the intervention subsequently), the only difference between the interventions in smartphone and control group was the experimental tools pupils were working with.

Research Questions

The theoretical framework elaborated above leads to the following main research questions (RQ) regarding affective and cognitive effects:

RQ1: Does using SETs lead to higher interest than using conventional experimental tools?
RQ2: Does using SETs lead to a higher state of curiosity than using conventional experimental tools?
RQ3: Is there a difference in learning achievement when using SETs compared to using conventional experimental tools?

It is understood that in all of the above comparisons, the smartphone group and the control group (conventional experimental tools) work on content identical experiments.

The Intervention

The intervention was comprised of experiments and associated instruction material for pupils, which used smartphones' acceleration sensors as measurement tools. As motivated above, the current study focused solely on the effects of trying to generate material context by introducing SETs in otherwise conventional settings. Harmonic mechanic oscillations, i.e., pendulum movements, were chosen as the topic of the intervention, in part because this is an important topic in the local syllabus (Ministerium für Bildung, Wissenschaft und Weiterbildung Rheinland Pfalz (MBWW) n.d.) as well as in the conceptual structure of physics (see, e.g., Baker and Blackburn 2005; Beech 2014; Matthews et al. 2005; Moshinsky and Smirnov 1996). Moreover, studying harmonic oscillations allows one to circumvent certain conceptual difficulties about acceleration and thus avoid confounding effects. Acceleration is difficult for pupils and often leads to misconceptions (see, e.g., Trowbridge and McDermott 1981). In conventional settings, acceleration is often not measured directly but calculated from measurements of displacement and time. If the smartphone group had to struggle with understanding acceleration–time graphs, while the control group was measuring displacement and time, the difficulties of the interventions would not be comparable. For harmonic oscillations, the displacement is proportional to the acceleration. In an investigation of the period of pendulum movements, the acceleration–time graphs on the smartphones can hence be interpreted as if they were displacement–time graphs: The period is the same for acceleration and displacement. To prevent misunderstandings, this approach was explained to pupils in the instruction material. With items about the representational competence in kinematics (see the "Test Instruments" section), we controlled that no misconceptions were caused or furthered.

Experiments and Instructional Material

During the intervention, each pupil conducted three experiments on the periodicity of a spring pendulum, a simple pendulum, and a coupled pendula (for experimental setups, see Fig. 1). The intervention was conducted as part of a curriculum unit about harmonic mechanic oscillations. Each participating teacher individually planned and taught the lessons before and after the intervention. After conferring with the teachers, it was agreed that all pupils would learn the formulas for calculating a spring or simple pendulum's period, but not conduct any experiments before the intervention took place. Pupils' prior knowledge can thus be a priori estimated as equal in both groups. Nevertheless, it was measured additionally in a pre-test (see the "Test Instruments" section). As pupils already knew the theory, the experiments about the spring and simple pendulum were used to apply and consolidate this prior knowledge. In the intervention, pupils investigated the influences of the displacement of the pendulum bob at the beginning of the oscillation, the mass of the pendulum bob, and the spring constant or the length of the string, respectively, on the period of the oscillation. The experiment regarding coupled

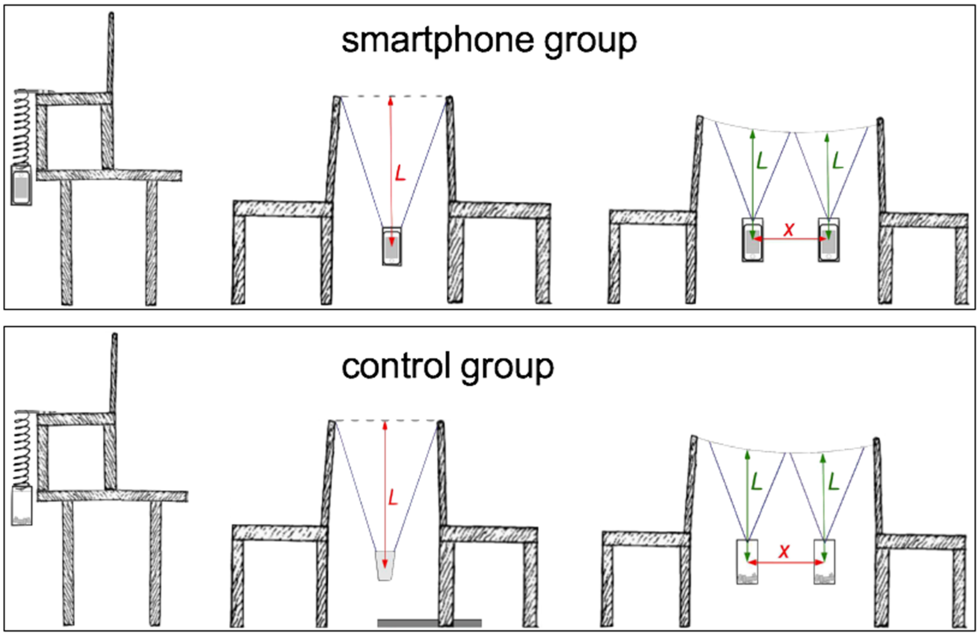

Fig. 1 Experimental set-ups of smartphone group (above) and control group (below): spring pendulum (left), simple pendulum (middle), coupled pendula (right)

pendula introduced a new topic: Pupils qualitatively examined the process of coupling and measured the influences of the distance of the two pendula and the tension of the connecting string on the coupling period.

The experimental set-ups of both groups were identical; the only difference being the use of smartphones as pendulum bobs while the control group was using screws. The app SPARKvue[1] displayed the acceleration data on the smartphones. It did not only show measurement data as real-time graphs but also offered the display of multiple measurements in one diagram, which simplified their comparison.

Pupils were provided with all necessary materials for conducting the experiments. All pupils in the smartphone group were given the same devices, regardless of whether they owned a smartphone. Where possible, everyday material was used to realize the experimental setups. All experiments could be done with any iOS or Android devices with the appropriate size and weight.

The instruction material was designed to be almost identically for both groups (see Online Resource 1 "Instruction Material"). For example, both groups were provided with worksheets that showed idealized graphs of the oscillation and with stopwatches for quantitative measurements. The control group compared the given, idealized graphs on the worksheets with their observation of the experiment and measured the period of the oscillations. On top of this, the pupils in the smartphone group produced several measurement graphs of different oscillations on the smartphone. They investigated the period of the oscillation not only with the measurements with the stopwatch but also by examining the acceleration–time graphs and interpreting them as if they were displacement–time graphs (see previous texts). They compared the different graphs on the smartphone with each other and with the given idealized graphs on the worksheets.

To summarize, there were multiple measures taken to avoid confounding variables and to concentrate the study solely on the potential effects of generating material context by using SETs: Choosing harmonic oscillations as topics of the intervention made it possible for both groups to study the same variables. Both groups were provided with graphs on the instruction material to ensure that all pupils were working with the same kinds of representations. As the experimental setups and associated instruction materials hardly differed, the cognitive load should be equal for both groups except for effects produced by using SETs. Moreover, the actual cognitive load perceived by pupils was measured to control for differences (see the "Test Instruments" section), either because of unintended inequalities in the instruction material or because of problems in handling the smartphones or a split attention effect (see the "Theoretical Background and Rationale" section).

Methods

Study Design

We conducted a quasi-experimental treatment group-control group study (see Fig. 2), which took place during pupils'

---

[1] The newest version of SPARKvue can be found under the following link: http://itunes.apple.com/de/app/sparkvue/id361907181 (PASCO Scientific 2015). After the current study was finished, the app was relaunched, which is why the interface of the version that was used in the study looked different from the version that can be obtained in the App store now.

regular physics lessons: Before the intervention, all pupils were tested for a total of 60 min. In the following four lessons (typically spread over one or two weeks), the intervention took place as described. In the next lesson, a post-test of 45 min was conducted; then, all pupils went back to their regular classes. After a period of 6 or 16 weeks (depending on the time of vacations), there was a follow-up test of 60 min.

Sample

In all, 245 pupils in 15 classes from 6 secondary schools in Rhineland-Palatinate, Germany, participated in the study (age range 14–19 years). In agreement with the participating teachers, ten classes were assigned to the smartphone group and five to the control group. If there were multiple participating physics classes in one school, at least one of them was assigned to the control group. The distribution of boys and girls was almost equal across treatment groups (SG 43% girls, CG 45% girls). Each group was taught by the same tutor.

Due to varying participation at different measurement times in the repeated measure design (see Fig. 2), sample size fluctuated slightly between measurement times, as given in Table 1. The numbers decrease for t_3 as some classes (unluckily, classes with an especially high percentage of girls) could not be tracked until the follow-up test 16 weeks later). For the validation of the test instruments, all available data of the respective measurement time were used. For the investigation of affective and cognitive effects, only data of pupils who were present for the pre-test, the post-test, and all lessons of the intervention were used: $N = 154$ pupils (60% boys, 40% girls, $M_{\text{age}} = 16.7$ years, age range 14–18 years).

Test Instruments

Motivational Variables

The test for measuring motivational variables was adapted from a well-tried and validated instrument (Kuhn 2010, respectively Kuhn and Müller 2014). For scale characteristics of the original scales, see Online Resource 2 “Origins of adapted test items with original number of items and reliability”, for item examples, see Table 2. The test used in the study combined the dependent variable interest (IN), and the control variables physics lessons’ perceived relation to reality (RR) and self-concept (SC). All items were related to physics classes in general, not only to the lessons of the intervention, i.e., IN measured individual interest, not situational interest (Hidi and Renninger 2006). All items were assessed using 6-point Likert scales. For conducting statistical analyses, the mean score of every item was transformed into a percentage, where 0% meant no affirmation of the statement and 100% meant a full affirmation of the statement. In addition to the original subscales of Kuhn, five items regarding pupils’ personal assessment of their teacher in the pre-test and their tutor in the post-test (assessment of teacher/tutor, AT) were adapted from Molz (2016). These items also used 6-point Likert scales, and their mean scores were transformed into percentages as shown above.

Curiosity

Regarding curiosity, there were two different variables adapted from two sources (for scale characteristics of the original scales, see Online Resource 2, “Origins of adapted test items with original number of items and reliability”, for item examples, see Table 3): In the pre-test, curiosity as a trait (CT) was measured, using seven items that were adapted from two well-tried and validated instruments (Litman and Spielberger 2003; Naylor 1981). For those items, a 4-point Likert scale was used and the mean scores were also transformed into percentages as seen above.

In the post-test, curiosity as a state related to the content of experiments (CSE) was measured with six items also adapted from Litman and Spielberger (2003) and Naylor (1981). All items used 6-point Likert scales. Again, the mean score was transformed into a percentage. For negated items the scale was inverted, so that 100% always stood for the highest evidence of curiosity.

Learning Achievement

To assess learning achievement (LA), a self-compiled concept test was conducted (example items see Table 4). It consisted of eight multiple-choice items in the pre-test and 12 in the post-test as well as five multiple true false items in the pre- and post-tests, which were each scored with 1 point for a right answer. Furthermore, the test contained three assertion reason tasks[2] in the pre-test and post-test, and five items regarding drawing or marking diagrams in the pre-test and six in the post-test. The free text and drawing items were categorized, rated, and accordingly scored as 0, 0.25, 0.5, 0.75, or 1 point. The total score was calculated as the sum of points divided by number of items; hence, it was always a percentage between 0% and 100%. In addition to the items designed to test learning achievement, ten multiple true false items, adapted from Klein et al. (2017), were used to assess representational competence in kinematics (KiRC). For scale characteristics of the original scale, see Online Resource 2, “Origins of adapted test items with original number of items and reliability”. Using this measurement, we could ensure that the interpretation of the acceleration–time graphs as displacement–time graphs

---

[2] Assertion reason task: multiple choice/true false item in combination with free text item asking for the reason behind the given answer (see, e.g., Williams 2007)

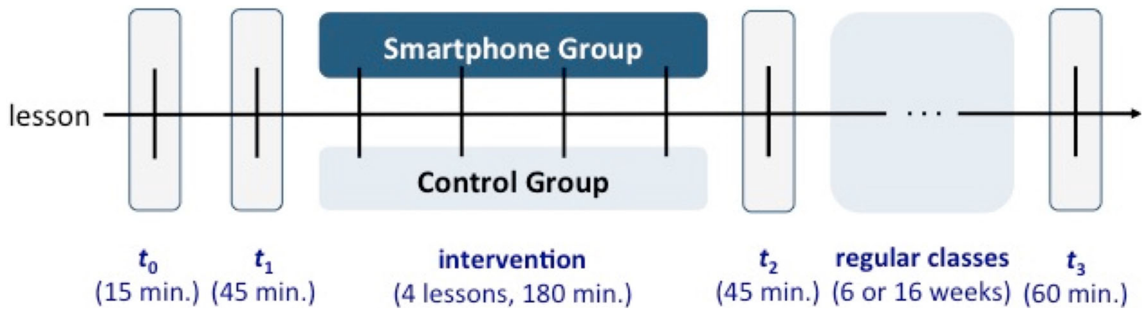

Fig. 2 Process of study. t_0 test of control variables, t_1 pre-test, t_2 post-test, t_3 follow-up test

(see the "The Intervention" section), did not lead to or further any misconceptions.

Control Variables

Three subscales were used to measure pupils' perceived cognitive load. The subscales for cognitive load related to the experiments (CLE) and cognitive load related to smartphones (CLS) were adapted from Kuhn (2010), Paas et al. (1994), and Chandler and Sweller (1991). For scale characteristics of the original scales, see Online Resource 2, "Origins of adapted test items with original number of items and reliability", for item examples, see Table 5. All items of the CLE and CLS scales used 6-point Likert scales. The CLS subscale was only used in the smartphone group. Furthermore, five items of the well-tried and validated NASA TLX instrument (Hart and Staveland 1988) were used, each with a 6-point bipolar scale. As in the other instruments, the scales of negated items were inverted and the mean scores transformed into percentages, so that 100% correlated the highest evidence of cognitive load.

In addition to the test instruments described previously, the following data were taken before the intervention to control confounding factors: experience with experiments (EE), possession and frequency of using smartphones (PUS), spatial ability (SA), and gender (G) (see Table 6). SA was assessed with a subtest of a standard test for cognitive abilities (Heller and Perleth 2000, nonverbal subtest 3, for scale characteristics of the original scale, see Online Resource 2, "Origins of adapted test items with original number of items and reliability").

Table 1 Sample sizes (245 students undergoing the intervention)

	t_0	t_1	t_2	t_3
N_{total}	223	225	203	172
N_{male}	123	126	109	126
N_{female}	99	99	80	42
M_{age}	16.7 years	16.7 years	16.6 years	17.0 years

Data Analyses

Preparation of Data

Unreliable information as distinct patterns in the answers to the Likert scales or unreadable answers were treated as missing values. For each variable, missing values were less than 5% of the data. Outliers were treated according to the procedure proposed by Tabachnik and Fidell (1996) (values with $|z| > 3.29$, i.e., 3 SDs off the average M, are replaced by $M \pm 3$ SD). For each variable, the percentage of outliers did not exceed 1%.

Analyses of Test Instruments

To confirm or modify intended subscales, principal component analyses were conducted. Furthermore, several item characteristics were calculated: the item difficulty P (for learning achievement) or the mean M (for rating scales), the discrimination index D, the corrected item-test-correlation r_{it}, and the reliability of the scale without a particular item $\alpha_{/\{i\}}$. As conventional critical values for psychometrically acceptable items, $D = 0.2$ and $r_{it} = 0.3$ were used (Ding and Beichner 2009). The scales were optimized by reliability analyses. All items that decreased the internal consistency Cronbach's α (as a measure of reliability) and whose characteristics were beyond the conventional critical values were deleted from the scale. We call these combinations of items optimized scales.

Analyses of Affective and Cognitive Effects

To gain as much information as possible out of gathered data, multiple-regression analyses were used when pre-conditions were fulfilled. Hierarchical regression analyses were used to make sure only as many variables as necessary went into the model to increase statistical power. In a hierarchical regression analysis, similar to a stepwise analysis, the regression model is extended by one predictor at the time. Only variables, which are explaining a significant amount of additional variance at the moment when they are considered, stay in the model. An advantage of the hierarchical regression in comparison to a

Table 2 Motivational variables and tutor/teacher assessment (sample items translated from German, more items in post-tests due to more time for testing, repeated items formulated equally in pre- and post-tests)

Subscale	Number of items	Example
Interest (IN)	Pre-test: 9 Post-test: 10	"In my leisure time, I engage in topics that are related to physics in excess of my homework."
Relation to reality (RR)	Pre-test: 5 Post-test: 8	"Physics classes are about things that are related to everyday life."
Self-concept (SC)	Pre-test: 8 Post-test: 8	"My achievement in physics is, in my opinion, good."
Personal assessment of teacher/tutor (AT)	Pre-test: 5 Post-test: 5	"Our teacher/tutor is/was motivating me."

stepwise regression is that the order of variables is not merely defined by statistics, but set because of theoretical reasons.

The following assumptions were tested for the applicability of multivariate regression analysis: no bias because of multivariate outliers, homoscedasticity, independence of the residuals, linearity, no autocorrelation, no multicollinear problems, normal distribution. If those assumptions were not met, analyses of variance (ANOVA) or covariance (ANCOVA) were used to analyze affective and cognitive effects. For these analyses, the following assumptions were tested: independence and normal distribution of residuals and homogeneity of residuals' variances, and additionally, for analyses of covariance homogeneity of regression slopes, adequate correlation between the covariate and the dependent variable, sufficient reliability of the covariate, and measurement of the covariate prior to the intervention.

Results

Analyses of Test Instruments

The principal component analysis confirmed the intended subscales of the test instrument for motivational variables (interest, IN; physics lessons' perceived relation to reality, RR; self-concept, SC). The test for learning achievement (LA) did not show any subscales; hence, the overall mean of all items is reported. In contrast to three intended separate subscales of cognitive load (CLE, CLS, TLX, see Table 5), all non-smartphone related items (CLE and TLX) formed one component, which was labeled general cognitive load (GCL). The second component consisted, as intended, of the items in the CLS scale.

Table 3 Curiosity variables (sample items translated from German)

Subscale	Number of items	Example
Curiosity as a trait (CT)	7	"Reviewing previous ideas is fun to me."
Curiosity as a state related to content of experiments (CSE)	6	"The experiments piqued my curiosity regarding the topic of oscillations."

The number of items (N_I), the reliability (Cronbach's α), the mean of discrimination indices $\overline{D}$, and the mean of item-test correlations $\overline{r_{it}}$ of the original scales and the scales with optimized Cronbach's α (items which decrease Cronbach's α were deleted to form optimized scales, see the "Data Analyses" section) can be found in Table 7. The mean of discrimination indices and the mean of corrected item-test correlations were within the conventional critical values ($D = 0.2$ and $r_{it} = 0.3$, see the "Data Analyses" section). The reliability exceeded $\alpha = 0.7$ for most subscales, at least in their optimized form. Only the reliability of representational competence (KiRC) did not reach this value.

Possession and frequency of using smartphones (PUS) were very high: 93% of the pupils owned a smartphone and they reported to use it often or very often ($M = 1.53$ on a scale from 1 to 6). Both variables were not used in the regression analyses or analyses of covariance to avoid ceiling effects.

Analyses of Affective and Cognitive Effects

Interest (RQ1)

Due to multicollinearity problems, interest (IN) could not be investigated by regression analysis. However, the assumptions for calculating ANCOVA (see the "Data Analyses" section) were met. The ANCOVA showed a significant small to mid-size effect between groups: $F\ (1, 149) = 6.11$, $p = 0.015$, $d = 0.40$ (adjusted group means, see Fig. 3 and Table 8). This effect could be specified using t tests for paired samples to investigate the difference between pre- and post-tests[3]: With a median split, the sample was divided into originally more ($IN_{\text{pre}} > 0.40$) and originally less interested pupils ($IN_{\text{pre}} \leq 0.40$). Originally less interested pupils in the smartphone

---

[3] For this analysis, the post-test mean is determined only from items that are also used in pre-test.

Table 4 Example items for the test instrument used to measure learning achievement (translated from German)

Item format	Example for items
Multiple choice item	In the picture, you see a wooden toy, which can swing up and down on a metal spring after displacement.
	 What is the influence of the mass of the wooden figure on the period of the oscillation of the toy? ☐ The **greater** the mass, the longer the period. ☐ The **smaller** the mass, the longer the period. ☐ The mass of the wooden figure has **no influence** on the period.
Multiple true false Items	The period of a simple pendulum is longer, ... a) ... if its **mass** is greater. ☐ True ☐ False b) ... if the **length of its string** is longer. ☐ True ☐ False
Assertion reason task	In reality, a pendulum that was set in motion once is oscillating **forever**. ☐ Right ☐ Wrong Justify your answer!
Drawing diagrams	**Draw** in the field below the diagram of an oscillation with a continuously **decreasing** amplitude and a **constant** period!

group profited most from the intervention: $t(41) = 2.24$, $p = 0.030$, $d = 0.23$ (group means, see Fig. 3). The pre–post-effects in all other subgroups (SG high interest, CG both subgroups) were not significant.

Curiosity (RQ2)

The conditions for hierarchical multiple-regression analysis were met, and it leads to the following model equation: $\hat{Y}(CSE)_i = 0.13 + 0.09\ (T_i) - 0.02\ (G_i) + 0.09\ (RR_i) + 0.06\ (SC_i) + 0.41\ (IN_i)$, with *CSE*, state of curiosity regarding the content of the experiments; *T*, treatment group; *G*, gender; *RR*, relation to reality; *SC*, self-concept; *IN*, interest. There was a significant small effect of the treatment group: $b^* = 0.21$, $t(150) = 3.19$, $p = 0.002$, $d = 0.25$. The only other significant predictor was the level of interest (IN, see Table 9).

Table 5 Subscales of the cognitive load test instrument (sample items translated from German)

Subscale	Number of items	Example
Cognitive load related to the experiments (CLE)	10	"I had to make an effort to solve the tasks of the experiments."
Cognitive load related to smartphones (CLS)	7	"I did not have any problems getting used to the app."
NASA TLX instrument (TLX)	5	"How successful were you in conducting the experiments in your own opinion? Perfect success—failure"

Learning Achievement (RQ3)

The conditions for hierarchical multiple-regression analysis were met, and it led to the following model equation: $\hat{Y}(LAPOST)_i = 0.26–0.02\ (T_i) + 0.01\ (G_i) + 0.06\ (EE_i) + 0.07\ (CT_i) + 0.09\ (SC_i) + 0.21\ (IN_i) + 0.14\ (KiRC_i) + 0.25\ (LAPRE_i)$, with *LAPOST*, learning achievement in post-test; *T*, treatment group; *G* gender; *EE* experience with experiments; *CT* curiosity as a trait; *SC* self-concept; *IN* interest; *KiRC* representational competence in kinematics; *LAPRE* learning achievement in pre-test. There was no significant effect of the treatment group: $b^* = -0.04$, $t(150) = -0.70$, $p = 0.482$. Significant predictors were the level of interest (IN), the representational competence in kinematics (KiRC), and the learning achievement (LA) in the pre-test (see

Table 6 Other control variables (sample items translated from German)

Variable	Number of items	Item format
Experience with experiments (EE)	2	"In my physics classes, our teacher shows us experiments: very often/often/rather often/rather rarely/rarely/very rarely" "In my physics classes, we do experiments ourselves: very often/often/rather often/rather rarely/rarely/very rarely"
Possession and frequency of using smartphones (PUS)	2	"I own an iPhone/other smartphone: yes/no" "I use the device: very often/often/rather often/rather rarely/rarely/very rarely"
Spatial abilities (SA)	15	"Each item shows how a quadratic gray piece of paper is folded and how holes are stamped into the folded paper. The task is to find out what the paper looks like when unfolded again."
Gender (G)	1	"Gender: male/female"

Table 10). Effect sizes regarding pre–post-differences were high: SG: $t(86) = 8.62$, $p < 0.001$, $d = 0.97$, CG: $t(66) = 9.39$, $p < 0.001$, $d = 1.01$ (for group means, see Fig. 4).

Control Variables

Unexpectedly, the ANCOVA showed no significant effect between groups for physics lessons' perceived relation to reality (RR): $F(1, 151) = 0.55$, $p = 0.459$ (adjusted group means, see Table 11).

Regarding assessment of tutor and cognitive load, no significant effects between treatment groups were found (see Table 12, Fig. 5). The cognitive load generated by using smartphones (CLS), which was only assessed in the smartphone group, was perceived as rather low (see Fig. 5).

Neither self-concept (SC), experience with experiments (EE), spatial abilities (SA), nor gender (G) had a significant influence on the analyses of interest (IN), curiosity (CSE), or learning achievement (LA) (see Tables 8, 9, and 10). In contradiction to our expectation (see the "Focus on Material Context in the Current Study" section), curiosity as a personality trait (CT) also had no significant effect on curiosity (CSE) after the intervention. The variable did not explain a significant amount of additional variance and was hence not included in the multiple-regression model (see Table 9). Nevertheless, curiosity as a trait (CT) did have an influence

Table 7 Test instruments: number of items (N_{I}), reliability (Cronbach's α), mean of discrimination indices $\overline{D}$, mean of item-test correlations $\overline{r_{\mathrm{it}}}$, N_{I} *of* post-test scales higher because of more testing time, N_{I} of optimized scores lower due to item deletion to reach optimal Cronbach's α

Scale		N_{I}	α	$\overline{D}$	$\overline{r_{\mathrm{it}}}$
Interest (IN)	Pre-test	9	0.92	0.58	0.70
	Post-test	10	0.96	0.57	0.78
Relation to reality (RR)	Pre-test	5	0.88	0.54	0.68
	Pre-test, optimized	4	0.90	0.62	0.78
	Post-test	8	0.94	0.52	0.74
Self-concept (SC)	Pre-test	8	0.90	0.62	0.72
	Post-test	8	0.92	0.64	0.80
Assessment of teacher (AT)	Pre-test	5	0.88	0.56	0.72
Assessment of tutor (AT)	Post-test	5	0.77	0.39	0.55
Curiosity as a trait (CT)		7	0.83	0.44	0.58
Curiosity as a state related to content of experiments (CSE)		6	0.91	0.54	0.74
Learning achievement (LA)	Pre-test	24	0.69	0.35	0.25
	Pre-test, optimized	16	0.78	0.46	0.36
	Post-test	31	0.83	0.43	0.34
	Post-test, optimized	27	0.85	0.47	0.38
Representational competence in kinematics (KiRC)	Pre-test	10	0.65	0.58	0.31
	Pre-test, optimized	8	0.67	0.67	0.36
	Post-test	10	0.67	0.59	0.33
General cognitive load (GCL)		15	0.89	0.36	0.56
General cognitive load (GCL), optimized		13	0.90	0.37	0.60

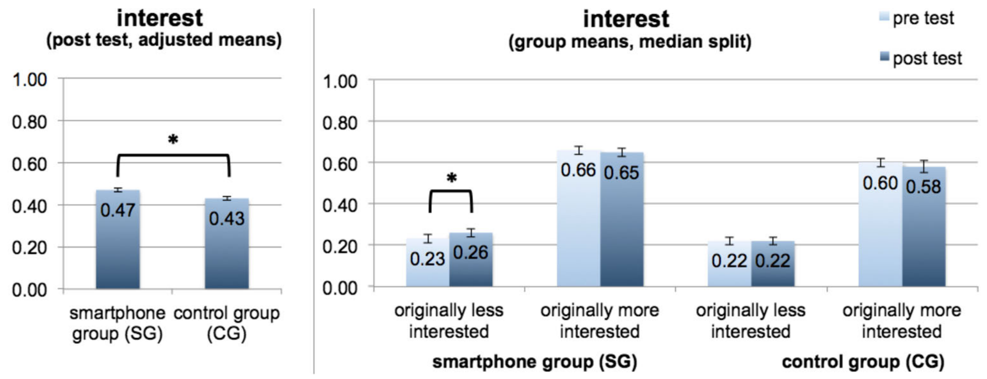

Fig. 3 Interest (IN): left: adjusted means and standard errors ($SE^*_{SG} = SE^*_{CG} = 0.01$); right: group means and standard errors pre- and post-tests ($SE_{SG} = SE_{CG} = 0.02$ for all means except post-test mean of originally more interested: $SE_{CG} = 0.03$; division of sample by median split: Mdn(IN) = 0.40)

on interest (IN) and was used as a covariate in the ANCOVA (see Table 8). For the multiple-regression analysis of learning achievement (LA), curiosity as a trait (CT) was included in the model, but did not have significant influence (see Table 10).

In addition to the control variables described above, a *t* test for paired samples was used on the pre- and post-test data of representational competence in kinematics (KiRC) to check whether misunderstandings were caused or deepened by interpreting acceleration-time-graphs as if they were displacement–time graphs (see the "The Intervention" section). There was no significant change of the representational competence over the course of the intervention in the smartphone group ($M_{\text{pre, SG}} = 0.65$; $M_{\text{post, SG}} = 0.64$; $t = 0.64$; $p = 0.525$).

Follow-Up Tests

Even though effect sizes for treatment effects were only small to midsize, follow-up tests were investigated. For this, similar models were used as for the post-tests. In contrast to the post-test models, all variables measured prior to the follow up test could now be used as predictors. For none of the variables, there was a significant effect of the treatment group.

Discussion

Interest in the smartphone group was significantly higher than that in the control group ($d = 0.40$, RQ1), even after a relatively short intervention (3 h "instructional episode" in the sense of Swarat et al. 2012). This is consistent with previous research about technology-based activities, concluding that they foster students' interest by connecting them with "real data" and providing a "sense of authenticity" (see Swarat et al. 2012 and literature cited therein). This is true even though this "sense of authenticity" was only provided by what we call a "material context" (see the "Theoretical Background and Rationale" section), the experimental tasks to be carried out being of a conventional kind, not explicitly related to a real-life context (in order to manipulate only one variable at a time).

Also in accord with the theory elaborated in the previous texts, curiosity regarding the content (!) of the experiments (and not regarding the device) was significantly more pronounced in the smartphone group than in the control group ($d = 0.25$, RQ2). This appears as a small effect, but it is noteworthy that the effect of the short-term intervention (i.e., "treatment" in the regression analysis) was almost half as large than that of the stable trait interest (prior to intervention) on curiosity. Apparently, consistent with the theoretical framework of curiosity above, finding information with SETs by conducting experiments raised pupils' confidence in being able to find a satisfactory solution for future problems and hence supported their curiosity.

Table 8 Adjusted means (M^*), standard errors (SE^*) and covariates with means (M) of interest (IN), sample size: $N_{SG} = 87$, $N_{CG} = 67$

Adjusted mean	Standard error	Covariates
$M^*_{SG} = 0.47$; $M^*_{CG} = 0.43$	$SE^*_{SG} = 0.01$; $SE^*_{CG} = 0.01$	Pre-test interest (IN), $M = 0.44$ Curiosity as a trait (CT), $M = 0.71$

Table 9 Hierarchical multiple-regression results: state of curiosity related to content of experiments (CSE), change in determination coefficient R^2, standardized regression coefficients b^*, t tests of the predictors with significance p and effect size f^2; characteristics of the whole model (R^2, adjusted R^2, standard error of estimation SEE, F test of whole model with significance p and effect size f^2); significant results in bold

Variable	Change in R^2	b^*	t	p	f^2
Treatment (T)	**0.056**	**0.21**	**3.19**	**0.002**	**0.04**
Gender (G)	0.057	− 0.36	− 0.51	0.608	< 0.01
Pre-test relation to reality (RR)	0.125	0.09	1.11	0.269	0.01
Pre-test self-concept (SC)	0.070	0.05	0.51	0.610	< 0.01
Pre-test interest (IN)	**0.069**	**0.45**	**4.03**	**< 0.001**	**0.08**
Model: $R^2 = 0.38$; adj. $R^2 = 0.36$; $SEE = 0.17$; $F = 17.66$; $p < 0.001$; $f^2 = 0.61$					

Table 10 Multiple regression: learning achievement in post-test (*LAPOST*), change in determination coefficient R^2, standardized regression coefficients b^*, t tests of the predictors with significance p and effect size f^2; characteristics of the whole model (R^2, adjusted R^2, standard error of estimation *SEE*, F test of whole model with significance p and effect size f^2); significant results in bold

Variable	Change in R^2	b^*	t	p	f^2
Treatment (T)	0.001	− 0.04	− 0.70	0.482	< 0.01
Gender (G)	0.059	0.03	0.45	0.654	< 0.01
Experience with experiments (EE)	0.044	0.09	1.36	0.176	< 0.01
Curiosity as a trait (CT)	0.066	0.07	1.02	0.307	< 0.01
Self-concept (SC)	0.147	0.11	1.13	0.262	< 0.01
Interest (IN)	**0.068**	**0.29**	**2.75**	**0.007**	**0.03**
Representational competence in kinematics (KiRC)	**0.043**	**0.18**	**2.26**	**0.026**	**0.02**
Learning achievement in pre-test (LAPRE)	**0.052**	**0.27**	**3.80**	**< 0.001**	**0.06**
Model: $R^2 = 0.48$; adj. $R^2 = 0.45$; $SEE = 0.13$; $F = 16.58$; $p < 0.001$; $f^2 = 0.92$					

It is noteworthy that neither interest nor curiosity was significantly dependent of pupils' prior knowledge. Moreover, there was no influence of self-concept, prior experience with experiments, spatial abilities, or gender on affective and cognitive effects. This means that by using SETs, the curiosity of pupils with both lower and higher initial experience and knowledge can be promoted equally well. On the one side, there were neither difficulties of conducting the experiments, nor was there a high cognitive load or novelty that would impede learning achievement, interest, or curiosity. On the other side, high-achieving pupils were not under-challenged or bored by the experiments, either. In addition, experimenting with SETs especially increased the interest of pupils less interested prior to the intervention (a small, but significant effect, $d = 0.25$ for the pre–post increase). This means that SETs might have the potential of closing the interest gap between more and less engaged pupils.

With respect to the cognitive variables (RQ3), significant learning gains with large pre–post effect sizes were found ($d_{SG} = 0.95$, $d_{CG} = 1.09$), showing high learning effects from the intervention both in the control and smartphone groups. However, no significant differences in the learning gains between the two groups were found. As both groups were working with almost identical tasks, involving the same variables and kinds of representations, cognitive activities were almost identical and no or only small differences could be expected from them. Moreover, treatment effects on learning achievement because of group differences in interest and curiosity are possible, but as both known meta-analytic correlations with achievement ($r \approx 0.3$, von Stumm et al. 2011; Uguroglu and Walberg 1979, respectively) and the primary effects found ($d = 0.4$ and $d = 0.25$, see previous texts) are not large, no or small effects due to this indirect reason are probable, too. For these reasons, it is not surprising that no cognitive advantages for the smartphone group could be found in the data. However, it is worth of pointing out that there were also no *disadvantages* in learning achievement because of using SETs either. Despite concerns about potentially increased cognitive load by this kind of instructional technology, no differences in cognitive load were found, confirming that there were no additional difficulties because of problems in handling the smartphones. Instead, pupils handled the smartphones and the app with ease, which can be seen by the low measure of cognitive load related to the devices. As

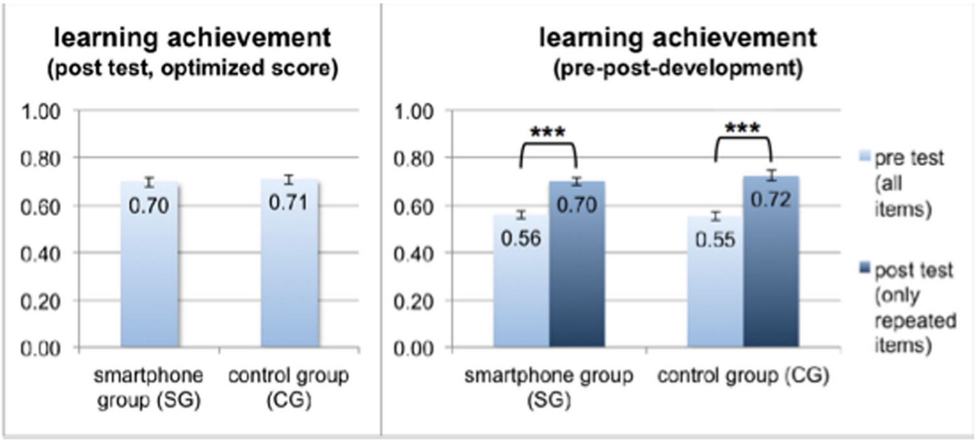

Fig. 4 Learning achievement (LA), group means, and standard errors: left: post-test, optimized score (repeated items from pretest and additional items), $SE_{SG} = 0.02$, $SE_{CG} = 0.02$; right: pre-test: all items, $SE_{SG} = 0.01$, $SE_{CG} = 0.02$, post-test: all items which were repeated from pre-test, $SE_{SG} = 0.02$, $SE_{CG} = 0.02$

Table 11 Adjusted means (M^*), standard errors (SE^*), and covariates with means (M) of relation to reality (RR), sample size: $N_{SG} = 87$, $N_{CG} = 67$

Adjusted mean	Standard error	Covariates
$M^*_{SG} = 0.52$; $M^*_{CG} = 0.51$	$SE^*_{SG} = 0.01$; $SE^*_{CG} = 0.02$	Pre-test relation to reality (RR), $M = 0.51$

there were no disadvantages in learning achievement for the smartphone group, we can conclude that pupils were not distracted by using SETs. This might be due to the fact that using SETs was integrated effectively into the instruction material and the teaching contexts. As opposed to the study of Tossell et al. (2014), pupils were not just provided with smartphones, but also with specific tasks. Another factor that might have helped pupils to concentrate on these tasks is their academic level: Beland and Murphy (2015) found that especially lower-achieving pupils were distracted by smartphones. Our sample consisted of pupils of the German "Gymnasium," which is the secondary school with the highest academic level, hence less vulnerable to distraction.

There were no treatment effects in the follow-up tests, which is not surprising, as the size of affective or cognitive effects between treatment groups was not large (as expected).

Regarding physics lessons' perceived relation to reality, there was no group difference, although the data shows the expected increase in interest and curiosity. Thus while the "material" context of SETs can foster interest and curiosity, it cannot alone strengthen the perceived relation to reality. Our interpretation of this finding is that for this to happen the experimental tasks have to be related to a real-life context, thus providing additionally a "topical" context (in the sense of the "Theoretical Background and Rationale" section), a dimension deliberately not considered within the research questions of the present contribution.

We now turn to a potential novelty effect. Short interventions as ours (180 min) especially if they are using modern technological devices run the risk of reporting a mere effect of novelty: The intervention might be increasing affective variables, just because the pupils experience something new and different from their usual physics classes. In our data, there are several reasons that lead us to conclude that the observed effects are not due to novelty, but to the intended advantages of SETs. While novelty is a primary factor of curiosity (see Berlyne 1978), this is not true for interest. For this reason, a novelty effect should lead to an effect on curiosity (where there is a theoretical reason) that exceeds the effect size of interest (where there is not), which is the opposite of the results found here. Moreover, effect sizes of novelty effects reported by meta-analysis (Adair et al. 1989) are not significantly different from zero, whereas for interest and curiosity we found significant effects (effect sizes 0.4 and 0.25, respectively); one would thus need a theoretical argument as to why in the given study novelty would have above-average, non-zero effects. Regarding the novelty generated by using SETs, another study (on the use of iPads in science learning, Miller et al. 2013) concluded that "the common presence of these personal electronics in student's lives today would likely reduce a novel effect in today's classroom" (Miller et al. 2013, p. 903). As in our sample, 93% of pupils stated to own a smartphone and to use it often or very often, it is much more likely, that the observed effects are not due to novelty, but to SETs' provision of material context respectively relevant sources for obtaining information.

In a study as ours where no big effects can be expected, it is even more important that multiple measures are taken to avoid confounding variables and make sure that the observed effects are only due to the intended difference between groups. The data that provide additional evidence against several confounding effects are reported here: The instruction material was identical in both groups, aside from the different measurement tools. By choosing harmonic oscillations as the topic of the intervention it was possible for both the smartphone and control group to study the same variables. As representational competence (KiRC results, Klein et al. 2017) stayed constant over the course of the intervention with smartphones, there was no evidence for an increased possible confusion of displacement and acceleration graphs due the measurement principle of the latter (see the "The Intervention" section). By providing both groups with graphs on the instruction material we ensured that all pupils were working with the same kinds of representations. The same tutor taught both groups and was in turn assessed by the pupils, with no significant difference found (no tutor bias). Moreover, as mentioned previously, there were no between-group differences in the perceived general cognitive load confirming that there were no additional difficulties because of inequalities in the instruction material. Negative effects of a potential distraction could not be found (see previous texts).

Table 12 Assessment of tutor (AT) and cognitive load (GCL/CLS): group means M, standard errors SE, t tests and their significance p, sample size: $N_{SG} = 87$, $N_{CG} = 67$ (cognitive load related to smartphones only assessed in smartphone group)

Scale	Group means	t test
Assessment of tutor (AT)	$M_{SG} = 0.74$, $SE_{SG} = 0.02$ $M_{CG} = 0.71$, $SE_{CG} = 0.02$	$t(152) = -0.97$, $p = 0.333$
General cognitive load (GCL)	$M_{SG} = 0.31$, $SE_{SG} = 0.02$ $M_{CG} = 0.31$, $SE_{CG} = 0.02$	$t(152) = 0.13$, $p = 0.895$
Cognitive load related to smartphones (CLS)	$M_{SG} = 0.14$, $SE_{SG} = 0.01$	–

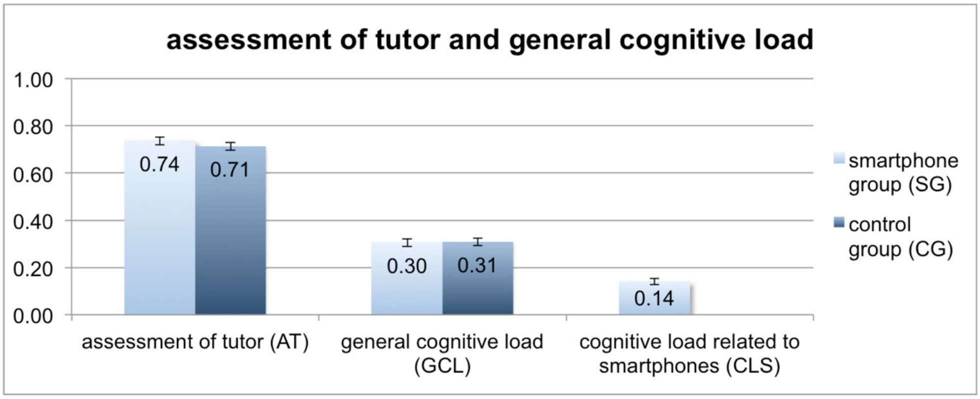

Fig. 5 Group means and standard errors of assessment of tutor (AT), general cognitive load (GCL), and cognitive load related to smartphones (CLS, only assessed in smartphone group)

Conclusions, Limitations, and Outlook

In the present study, SETs were used as mobile pocket-labs in an attempt to generate relevant material context in the sense of a relation to pupils' everyday life. The devices were used within standard, well-tried experimental setups involving pendulum movements in regular physics classes of the upper secondary level; the internal acceleration sensors of the devices provided pupils with acceleration and displacement data as graphs. Besides the already known practical and experimental advantages of SETs, we can conclude the following statements from our findings: (a) The use of SETs significantly raised pupils' interest regarding their physics classes in general and curiosity regarding the content of the experiments. Although effect sizes are small to medium, they occur after a relatively short intervention (3 h). (b) None of the often apprehended disadvantages (see, e.g., Beland and Murphy 2015; van Bruggen et al. 2002; Tossell et al. 2014) were found in the present study. Pupils were not over-challenged by the application of the new media: There was no difference in the perceived cognitive load between the smartphone and control group, and the cognitive load due to smartphones was perceived as rather small altogether. Furthermore, learning achievement in the smartphone group was as high as in the control group, making a potential distraction effect unlikely. (c) Findings of positive effects and the absence of negative ones do not depend on various covariates; in particular, there was no dependence on gender, self-concept, or experimental experience. (d) A potential of SETs that emerges from the study is to promote interest especially in pupils that were less interested at the beginning of the study. Together with the independence on initial knowledge and experience, this means that in a setting like the one presented in this contribution, SETs offer the chance to inspire especially less interested and lower achieving pupils to engage in science experiments without the demanding preconditions of experimental experience. This correlates with the work of Miller et al. (2013), who found similarly encouraging results for the use of science notebooks with tablet computers for pupils with intellectual disabilities.

The study at hand contributes to the discussion about the effective integration of technology in general (Swarat et al. 2012) and the use of SETs in science classes in particular. Similar to the now decades long discussion about using computers in schools, there is a lively debate (see, e.g., Barkham and Moss 2012; CBC News 2015; Jeffreys 2015; Mathews 2015) about the potential advantages and disadvantages of smartphones. In contrast to the older discussion related to computers, which can be supported by empirical results (Organization for Economic Cooperation and Development [OECD] 2015), in the case of smartphones empirical underpinnings of the discussion are found wanting. Problems that are generated by smartphones in everyday life and especially in school can surely not be denied and have to be considered and solved on a societal level. But as life without smartphones is unimaginable for most youths nowadays, the part of science education can only be, and has to be, to show new, meaningful possibilities for the application of smartphones inside and outside of school.

As there is empirical evidence for using modern instructional technology to support less favored learners in science classrooms, investigating this potential appears to be an interesting research question for the future. As our study was limited to pupils with a high academic level (German "Gymnasium") and to one age group (age range 14–19 years), it would be especially interesting to take a closer look at the affective and cognitive effects of using SETs at various ages and academic levels.

Moreover, further studies are necessary on the use of SETs and similar devices inside and especially outside the classroom for investigation of true real world phenomena and questions (the present study was deliberately restricted to the material context as explained in the theory section). Life-world contexts can be more authentic but usually are also far more open and complex. So, there might be stronger effects on motivation, but effects on learning and transfer remain to be studied.

In addition to the restriction to the material context, the experimental task of studying the influences of several variables only on the period of an oscillation allowed an optimal comparability of the treatment groups. Investigations regarding the affective and cognitive effects of examining the acceleration or other physical variables (using integrated sensors as, e.g., microphone, camera, or magnetic flux density sensor) would be necessary to obtain a more complete view of the potential of SETs in physics education.

Finally, future studies could exploit SETs still more fully as cognitive tools by integrating the possibilities of using different representations or real-time feedback. First steps along these lines have been taken (Klein et al. 2015; Kuhn and Vogt 2015; Mazzella and Testa 2016), but much remains to do in this innovative, but under-researched field.

Acknowledgements Generous financial support by the "Wilfried-und-Ingrid-Kuhn-Stiftung" for the doctoral thesis of Katrin Hochberg is gratefully acknowledged. The current paper is based largely on this thesis.

Funding This study was funded by the "Wilfried-und-Ingrid-Kuhn-Stiftung".

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest Katrin Hochberg declares that she has no conflict of interest. Jochen Kuhn declares that he has no conflict of interest. Andreas Müller declares that he has no conflict of interest.

References

Adair, J. G., Sharpe, D., & Huynh, C. L. (1989). Hawthorne control procedures in educational experiments: a reconsideration of their use and effectiveness. *Review of Educational Research, 59*(2), 215–228.

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learn Instr, 16*(3), 183–198.

Alrasheedi, M., Capretz, L. F., & Raza, A. (2015). A systematic review of the critical factors for success of mobile learning in higher education (university students' perspective). *J Educ Comput Res, 52*(2), 257–276.

Arnone, M. P., Small, R. V., Chauncey, S. A., & McKenna, H. P. (2011). Curiosity, interest and engagement in technology-pervasive learning environments: a new research agenda. *Educ Technol Res Dev, 59*(2), 181–198.

Bahtaji, M. A. A. (2015). Improving transfer of learning through designed context-based instructional materials. *European Journal of Science and Mathematics Education, 3*(3), 265–274.

Baker, G. L., & Blackburn, J. A. (2005). *The pendulum. A case study in physics*. New York: Oxford University Press.

Barkham, P., & Moss, S. (2012). Should mobile phones be banned from schools? The Guardian. https://www.theguardian.com/education/2012/nov/27/should-mobiles-be-banned-schools. Accessed 21 March 2017.

Beech, M. (2014). *The pendulum paradigm*. Boca Raton, Florida: Brown Waler Press.

Beichner, R. J. (1996). The impact of video motion analysis on kinematics graph interpretation skills. *Am J Phys, 64*(10), 1272–1277.

Beland, L. P., & Murphy, R. (2015). Ill communication: technology, distraction & student performance. Centre for economic performance. http://cep.lse.ac.uk/pubs/download/dp1350.pdf. Accessed 21 March 2017.

Bennett, J., Lubben, F., & Hogarth, S. (2007). 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*(3), 347–370.

Berlyne, D. E. (1978). Curiosity and learning. *Motiv Emot, 2*(2), 97–175.

Brasell, H. (1987). The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. *J Res Sci Teach, 24*(4), 385–395.

Castro-Palacio, J. C., & Velázquez-Abad, L. (2013). Using a mobile phone acceleration sensor in physics experiments on free and damped harmonic oscillations. *Am J Phys, 81*(6), 472–475.

CBC News. (2015). Smartphones in the classroom: a teacher's dream or nightmare? CBC News. http://www.cbc.ca/news/technology/smartphones-in-the-classroom-a-teacher-s-dream-or-nightmare-1.3211652. Accessed 21 March 2017.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cogn Instr, 8*(4), 293–332.

Chevrier, J., Madani, L., Ledenmat, S., & Bsiesy, A. (2013). Teaching classical mechanics using smartphones. *Phys Teach, 51*(6), 376–377.

Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education, 110*, 51–63.

Crompton, H., Burke, D., Gregory, K. H., & Gräbe, C. (2016). The use of mobile learning in science: a systematic review. *J Sci Educ Technol, 25*(2), 149–160.

DeWitt, J., & Storksdieck, M. (2008). A short review of school field trips: key findings from the past and implications for the future. *Visitor Studies, 11*(2), 181–197.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics—Physics Education Research, 5*(2).

Falk, J., & Balling, J. (1979). *Setting a neglected variable in science education: investigations into outdoor field trips*. Edgewater, MD: Smithsonian Institution, Chesapeake Bay Center for Environment Studies.

Forinash, K., & Wisman, R. F. (2012). Smartphones as portable oscilloscopes for physics labs. *The Physics Teacher, 50*(4), 242–243.

Fried, C. B. (2008). In-class laptop use and its effects on student learning. *Computers & Education, 50*(3), 906–914.

Gilbert, J. K., Bulte, A. M. W., & Pilot, A. (2011). Concept development and transfer in context-based science education. *International Journal of Science Education, 33*(6), 817–837.

Greenslade Jr., T. B. (2016). Whistling tea kettles, train whistles, and organ pipes. *The Physics Teacher, 54*(9), 518–519.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M. & Elliot, A. J. (2000). Short-term and longterm consequences of achievement goals: Predicting interest and performance over time. *Journal of educational psychology, 92*(2), 316.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. In N. Meshkati & P. A. Hancock (Eds.), *Human Mental Workload* (Vol. 52, pp. 139–183). Elsevier.

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: a critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139–156.

Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hazel, E., Logan, P., & Gallagher, P. (2007). Equitable assessment of students in physics: importance of gender and language background. *International Journal of Science Education, 19*(4), 381–392.

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4.-12.Klassen, Revision (KFT 4-12+R)*. Göttingen: Hogrefe.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127.

Hirth, M., Kuhn, J., & Müller, A. (2015). Measurement of sound velocity made easy using harmonic resonant frequencies with everyday mobile technology. *The Physics Teacher, 53*(2), 120–121.

Hochberg, K., Gröber, S., Kuhn, J., & Müller, A. (2014). The spinning disc: studying radial acceleration and its damping process with smartphone acceleration sensors. *Physics Education, 49*(2), 137–140.

Hoffmann, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning and Instruction, 12*(4), 447–465.

Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65–E70.

Jeffreys, B. (2015). Can a smartphone be a tool for learning? BBC News. http://www.bbc.com/news/education-34389063. Accessed 21 March 2017.

Klein, P., Hirth, M., Gröber, S., Kuhn, J., & Müller, A. (2014). Classical experiments revisited: smartphones and tablet PCs as experimental tools in acoustics and optics. *Physics Education, 49*(4), 412–418.

Klein, P., Kuhn, J., Müller, A., & Gröber, S. (2015). Video analysis exercises in regular introductory mechanics physics courses: effects of conventional methods and possibilities of mobile devices. In A. Kauertz, H. Ludwig, A. Müller, J. Pretsch, & W. Schnotz (Eds.), *Multidisciplinary research on teaching and learning* (pp. 270–288). Basingstoke: Palgrave Macmillan.

Klein, P., Müller, A., & Kuhn, J. (2017). Assessment of representational competence in kinematics. *Physical Review Special Topics—Physics Education Research, 13*(1), 010132.

Koleza, E., & Pappas, J. (2008). The effect of motion analysis activities in a video-based laboratory in students' understanding of position, velocity and frames of reference. *International Journal of Mathematical Education in Science and Technology, 39*(6), 701–723.

Kost-Smith, L. E., Pollock, S. J., & Finkelstein, N. D. (2010). Gender disparities in second-semester college physics: the incremental effects of a "smog of bias". Physical Review Special Topics—Physics Education Research, 6(2), 020112.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learn Instr, 15*(5), 381–395.

Kuhn, J. (2010). *Authentische Aufgaben im theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht*. Wiesbaden: Vieweg+Teubner.

Kuhn, J., & Müller, A. (2014). Context-based science education by newspaper story problems: a study on motivation and learning effects. *Perspectives in Science, 2*(1–4), 5–21.

Kuhn, J., & Vogt, P. (2012). iPhysicsLabs (series), column editors' note. *The Physics Teacher, 50*, 372–373.

Kuhn, J., & Vogt, P. (2013). Smartphones as experimental tools: different methods to determine the gravitational acceleration in classroom physics by using everyday devices. *European Journal of Physics Education, 4*(1), 16–27.

Kuhn, J., & Vogt, P. (2015). Smartphones & Co. in physics education: effects of learning with new media experimental tools in acoustics. In W. Schnotz, A. Kauertz, H. Ludwig, A. Müller, & J. Pretsch (Eds.), *Multidisciplinary research on teaching and learning* (pp. 253–269). Basingstoke: Palgrave Macmillan UK.

Kuhn, J., Molz, A., Gröber, S., & Frübis, J. (2014). iRadioactivity—possibilities and limitations for using smartphones and tablet PCs as radioactive counters. *The Physics Teacher, 52*(6), 351–356.

Lenhart, A. (2015). Teens, social media, and technology overview 2015. PewResearchCenter. http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/#. Accessed 21 March 2017.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*(1), 75–86.

Main, S., & ORourke, J. (2011). "New directions for traditional lessons": can handheld game consoles enhance mental mathematics skills? *Australian Journal of Teacher Eduation, 36*(2), 43–55.

Mathews, J. (2015). Are smartphones dumbing down school, or are they vital learning tools? The Washington Post. https://www.washingtonpost.com/local/education/are-smartphones-dumbing-down-school-or-are-they-vital-learning-tools/2015/10/25/3e278ac8-7a27-11e5-b9c1-f03c48c96ac2_story.html?utm_term=.736cb682baa4. Accessed 21 March 2017.

Matthews, M. R., Gauld, C. F., & Stinner, A. (2005). *The pendulum*. Dordrecht, The Nederlands: Springer Science & Business Media.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educ Psychol, 38*(1), 43–52.

Mazzella, A., & Testa, I. (2016). An investigation into the effectiveness of smartphone experiments on students' conceptual knowledge about acceleration. *Phys Educ, 51*(5), 055010.

Miller, B. T., Krockover, G. H., & Doughty, T. (2013). Using iPads to teach inquiry science to students with a moderate to severe intellectual disability: a pilot study. *Journal of Research in Science Teaching, 50*(8), 887–911.

Ministerium für Bildung, Wissenschaft und Weiterbildung Rheinland Pfalz [MBWW]. (n.d.). Lehrplan Physik: Gymnasium Sek. II. Bildungsserver. https://lehrplaene.bildung-rp.de/no-cache/gehezu/startseite.html?tx_pitsdownloadcenter_pitsdownloadcenter%5Bcontroller%5D=Download&tx_pitsdownloadcenter_pitsdownloadcenter%5Baction%5D=forceDownload&tx_pitsdownloadcenter_pitsdownloadcenter%5Bfileid%5D=NzQ0NDc%3D. Accessed 21 March 2017.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology* 85(3), 424–436.

Molz, A. (2016). *Verbindung von Schülerlabor und Schulunterricht – Auswirkungen auf Motivation und Kognition im Fach Physik*. München: Verlag Dr. Hut.

Monteiro, M., Cabeza, C., & Marti, A. C. (2014). Exploring phase space using smartphone acceleration and rotation sensors simultaneously. *Eur J Phys, 35*(4), 045013.

Monteiro, M., Vogt, P., Stari, C., Cabeza, C., & Marti, A. C. (2016). Exploring the atmosphere using smartphones. *Phys Teach, 54*(5), 308–309.

Moreno, R. (2005). Instructional technology—promise and pitfalls. In L. M. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education* (pp. 1–19). Greenwich, Conn.: IAP.

Moshinsky, M., & Smirnov, Y. F. (1996). *The harmonic oscillator in modern physics* (Vol. 9). Amsterdam: Harwood Academic Publishers.

Müller, R. (2006). Physik in interessanten Kontexten. Handreichung für die Unterrichtsentwicklung. https://www.tu-braunschweig.de/Medien-DB/ifdn-physik/physik-in-interessanten-kontexten-rmueller.pdf. Physik im Kontext. Accessed 21 March 2017.

Müller, A., Vogt, P., Kuhn, J., & Müller, M. (2015). Cracking knuckles—a smartphone inquiry on bioacoustics. *The Physics Teacher, 53*(5), 307–308.

Müller, A., Hirth, M., & Kuhn, J. (2016). Tunnel pressure waves—a smartphone inquiry on rail travel. *The Physics Teacher, 54*(2), 118–119.

National Science Foundation. (1983). *Educating Americans for the 21st century: report of the National Science Board on pre-college education in mathematics, science and technology*. Washington, DC: National Science Foundation.

Naylor, F. D. (1981). A state-trait curiosity inventory. *Australian Psychologist, 16*(2), 172–183.

Newhouse, P., & Rennie, L. (2001). A longitudinal study of the use of student-owned portable computers in a secondary school. *Computers & Education, 36*(3), 223–243.

Organisation for Economic Co-operation and Development [OECD]. (2007). *PISA 2006* (Vol. 2: Data). Pisa: OECD Publishing.

Organisation for Economic Co-operation and Development [OECD]. (2015). *Students, computers and learning: making the connection*. Pisa: OECD Publishing.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079.

Paas, F. G. W. C., van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419–430.

Parolin, S. O., & Pezzi, G. (2013). Smartphone-aided measurements of the speed of sound in different gaseous mixtures. *The Physics Teacher, 51*(8), 508–509.

PASCO Scientific. (2015). SPARKvue (Version 2.3.2). PASCO Scientific. https://itunes.apple.com/de/app/sparkvue/id361907181?mt=8. Accessed 21 March 2017.

Pimmer, C., Mateescu, M., & Gröhbiel, U. (2016). Mobile and ubiquitous learning in higher education settings. A systematic review of empirical studies. *Computers in Human Behavior, 63*, 490–501.

Ratcliffe, M., & Millar, R. (2009). Teaching for understanding of science in context: evidence from the pilot trials of the twenty first century science courses. *Journal of Research in Science Teaching, 46*(8), 945–959.

Reid, N., & Skryabina, E. A. (2002). Attitudes towards physics. *Research in Science & Technological Education, 20*(1), 67–81.

Sans, J. A., Manjón, F. J., Pereira, A. L. J., Gomez-Tejedor, J. A., & Monsoriu, J. A. (2013). Oscillations studied with the smartphone ambient light sensor. *European Journal of Physics, 34*(6), 1349–1354.

Shakur, A., & Sinatra, T. (2013). Angular momentum. *The Physics Teacher, 51*(9), 564–565.

Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage handbook of eLearning research*. London: Sage.

Silva, N. (2012). Magnetic field sensor. *The Physics Teacher, 50*(6), 372–373.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: understanding student interest in school science. *Journal of Research in Science Teaching, 49*(4), 515–537.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Taasoobshirazi, G., & Carr, M. (2008). A review and critique of context-based physics instruction and assessment. *Educational Research Review, 3*(2), 155–167.

Tabachnik, B. G., & Fidell, L. S. (1996). *Using multivaraite statistics*. New York: Harper Collins College Publishers.

Tho, S. W., Chan, K. W., & Yeung, Y. Y. (2015). Technology-enhanced physics programme for community-based science learning: innovative design and programme evaluation in a theme park. *Journal of Science Education and Technology, 24*(5), 580–594.

Thoms, L.-J., Colicchia, G., & Girwidz, R. (2013). Color reproduction with a smartphone. *The Physics Teacher, 51*(7), 440–441.

Thornton, R. K., & Sokoloff, D. R. (1990). Learning motion concepts using real-time microcomputer-based laboratory tools. *American Journal of Physics, 58*(9), 858–867.

Tornaría, F., Monteiro, M., & Marti, A. C. (2014). Understanding coffee spills using a smartphone. *The Physics Teacher, 52*(8), 502–503.

Tossell, C. C., Kortum, P., Shepard, C., Rahmati, A., & Zhong, L. (2014). You can lead a horse to water but you cannot make him learn: smartphone use in higher education. *British Journal of Educational Technology, 46*(4), 713–724.

Trowbridge, D. E., & McDermott, L. C. (1981). Investigation of student understanding of the concept of acceleration in one dimension. *American Journal of Physics, 49*(3), 242–253.

Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: a quantitative synthesis. *American Educational Research Journal, 16*(4), 375–389.

van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learn Instr, 12*(1), 121–138.

Vogt, P., Kasper, L., & Burde, J. (2015). The sound of church bells: tracking down the secret of a traditional arts and crafts trade. *The Physics Teacher, 53*(7), 438–439.

von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science, 6*(6), 574–588.

Wild, E., Hofer, M., & Pekrun, R. (2001). Psychologie des Lerners. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie - Ein Lehrbuch* (pp. 207–270). Weinheim: Beltz Psychologie Verlags Union.

Williams, J. B. (2007). Assertion-reason multiple-choice testing as a tool for deep learning: a qualitative analysis. *Assessment & Evaluation in Higher Education, 31*(3), 287–301.

Wu, W. H., Wu, Y. C. J., Chen, C. Y., Kao, H. Y., Lin, C. H., & Huang, S. H. (2012). Review of trends from mobile learning studies: a meta-analysis. *Computers & Education, 59*(2), 817–827.