ORIGINAL PAPER

Investigating Dynamic Visualizations of Multiple Representations Using Mobile Video Analysis in Physics Lessons

Effects on Emotion, Cognitive Load and Conceptual Understanding

Sebastian Becker[1] · **Pascal Klein[2]** · **Alexander Gößling[3]** · **Jochen Kuhn[1]**

Abstract
This contribution presents the results of a replication study on the learning effect of tablet-supported video analysis compared to traditional teaching sequences using non-digital experimental materials in the subject areas of uniform and accelerated motion in high school physics lessons. In addition to the replication of the preliminary study results recently published in this journal (Becker et al 2018, 2019), the investigation of the effect on the cognitive load as well as the emotional state of the students is another focal point. Compared to the preliminary study, the sample size was significantly increased from $N = 109$ to $N = 294$. The individual effects of the preliminary study could be replicated in this way. For both topics, a significant reduction of extraneous cognitive load and a positive effect on intervention-induced emotions could be demonstrated. Moreover, the theoretically founded causal relationship between emotion, cognitive load, and learning achievement could be empirically verified by means of structural equation modeling.

✉ Sebastian Becker
s.becker@physik.uni-kl.de

Pascal Klein
pascal.klein@uni-goettingen.de

Alexander Gößling
Alexander.Goessling@gmx.de

Jochen Kuhn
kuhn@physik.uni-kl.de

[1] Physics/Physics Education Research Group, Technische Universität Kaiserslautern, Kaiserslautern, Germany

[2] Faculty of Physics/Physics Education Research, Georg-August-University Göttingen, Göttingen, Germany

[3] Institute for School, Educational and Professional Science, Bielefeld, Germany

Untersuchung dynamischer Visualisierungen von multiplen Repräsentationen mittels mobiler Videoanalyse im Physikunterricht

Wirkungen auf Emotionen, kognitive Belastung und Konzeptverständnis

Zusammenfassung

In diesem Beitrag werden die Ergebnisse einer Replikationsstudie zu lernwirksamen Effekten der Tablet-PC-gestützten Videoanalyse im Vergleich zu traditionellen Unterrichtssequenzen unter Verwendung nicht-digitaler Experimentiermittel in den Themenbereichen gleichförmige und gleichmäßig beschleunigte Bewegung des Physikunterrichts der Sekundarstufe 2 vorgestellt. Neben der Replikation der jüngst in dieser Zeitschrift veröffentlichten Vorstudienergebnisse (Becker et al 2018, 2019) bildet die Untersuchung der Wirkung auf die kognitive Belastung sowie den emotionalen Zustand der Schülerinnen und Schüler einen weiteren Schwerpunkt. Im Vergleich zur Vorstudie wurde der Stichprobenumfang von $N = 109$ auf $N = 294$ deutlich erhöht. Einzelne Effekte der Vorstudie konnten auf diese Weise repliziert werden. Für beide Themengebiete konnte weiterhin eine signifikante Reduktion der extrinsischen kognitiven Belastung sowie eine positive Wirkung auf die interventions-induzierten Emotionen nachgewiesen werden. Darüber hinaus konnte der theoretisch fundierte kausale Zusammenhang zwischen Emotion, kognitiver Belastung und Lernperformanz mittels Strukturgleichungsmodellierung empirisch verifiziert werden.

Introduction

Although physical experiments play a key role in learning about science (e.g., Haury and Rillero 1994; Haagen-Schützenhöfer and Joham 2018), the learning-promoting potential offered by this learning activity is not sufficiently exploited (e.g., Woolnough 1979; Volkwyn et al 2008; Husnaini and Chen 2019; Kapici et al 2019). The causes of this discrepancy are manifold, as the learning objectives associated with an experimental learning activity can also be diverse (Hart et al 2000). Johnstone and Wham (1982), for example, argue that students in an experiment-based learning situation are cognitively overloaded by the amount of information they have to keep mentally available. Against this background, technologies for the digital acquisition and visualization of measurement data offer novel possibilities to support learners in experimenting. In contrast, studies on integrating such digital learning tools in experiment-based learning processes are still rare, especially for real-life teaching scenarios (Oliveira et al 2019). In this context, Zydney and Warner (2016) point out the need for empirical studies in order to better coordinate underlying theories and results on learning effects. Due to the increasing use of mobile technologies both in everyday life and in the education sector, Mutlu-Bayraktar et al (2019) recommend that the influence of mobile technologies on cognitive load in teaching-learning situations should also be systematically investigated in empirical studies. In this respect, initial studies have shown positive effects of using mobile devices to enhance inquiry-based experimental learning with multiple representations on conceptual understanding (e.g., Becker et al 2018, 2019, 2020; Klein et al 2018; Kuhn and Vogt 2015; Hochberg et al 2020) and motivation (e.g., Hochberg et al 2018).

In our preliminary study on the learning effectiveness of tablet-supported video analysis compared to traditional teaching sequences with non-digital experimental equipment, positive effects for the promotion of conceptual understanding for use in mechanics lessons have already been demonstrated (Becker et al 2018, 2019). As a result of the small sample size of this study (see Table 1), a generalization of the study findings is questionable. In view of the need for an evidence-based approach to pedagogical actions, however, the generalizability of scientific findings is important for their transfer to school practice. Furthermore, the interpretation of the results regarding the learning effectiveness remained hypothetical. Referring to the Cognitive Load Theory (see theoretical background), we argued that the use of the digital learning tool in a multirepresentational learning environment leads to a reduction of the extraneous cognitive load, which, in turn, results in an increased learning performance. In this view, load reduction leads to more free cognitive resources that are available to learners for active knowledge construction, which should increase the learning outcome. This is particularly true for multirepresentational learning environments, in which the complexity of information presentation can place an additional load on learners (van Meter et al 2020; Seufert 2003; de Jong et al 1998).

However, this hypothetical effect could not be empirically verified due to a methodological deficit of the previous study since the instrument used to determine the cognitive load did not allow a differentiation between the different types of load. Moreover, no affective variables were collected in the preliminary study that could have indirectly influenced learning performance. The analysis methods used also did not take into account the hierarchical data structure, so that teacher or school effects were not integrated into the analysis of learning outcomes, which could have led to in-

Table 1 Common and different methodological features of preliminary and replication study

	Preliminary study	Replication study
Sample		
Number of cases	109	294
Grade level	11	11
School year	2016/17	2017/18
Federal states	RLP, NRW	
Subject areas	Uniform motion, Accelerated motion	
Target variables		
Performance	Multiple choice concept test	
Cognitive load	NASA-TLX	CLS
Emotion	–	5-item scale
Teacher behavior	4-item scale	
Data analysis		
Performance	ANCOVA	Multilevel Regression Analysis
Cognitive load	Mann-Whitney U test	Multilevel Regression Analysis
Emotion	–	Multilevel Regression Analysis
Teacher behavior	Correlation Analysis	Structural Equation Analysis
Empirical causal analysis	–	Structural Equation Analysis

correct significance findings. To remedy these methodological deficits and to strengthen the statistical power, a replication study was conducted with the following objectives.

- Replication of learning-promoting effects of the preliminary study with increased case numbers
- Empirical verification of the hypothetical cause-effect relationship between extraneous cognitive load and learning performance
- Quantifying the influence from selected affective variables
- Consideration of teacher and school effects in the analysis of learning achievement

In order to show the reproducibility of the effects of the preliminary study, the replication study was carried out under exactly the same experimental conditions. In particular, the study design, the preparation of the participating teachers, the timing and structure of the teaching sequences, the design of the learning environments, the achievement tests, and the learning material were adopted unchanged from the preliminary study. The target population was recruited from the same grade level, and the intervention was carried out in a comparable period of the school year. An overview of the most important methodological features of the preliminary and replication study is given in Table 1.

In general, replication studies contribute significantly to empirical research by confirming previous results or re-

moving limitations of previous studies (Makel and Plucker 2014). They are particularly important for empirical educational research because of the limited control options in field studies (Lindsay and Ehrenberg 1993; Guilford 1982). Although replication studies can promote the transfer of knowledge gained into classroom practice and ideally accelerate this process, they are rare (Makel and Plucker 2014). Rost and Bienefeld (2019) see replication studies as an "irreplaceable means of testing and safeguarding scientific knowledge," which can "do what statistical significance tests and effect size calculations cannot do" (p. 9).

Against this background, this work aims to provide a counterpoint to the lack of replication studies in empirical educational research and in this way contribute to the generalizability of the effects found in the preliminary study on the one hand, and to empirically verify theoretically hypothesized causes of the learning-promoting effect on the other.

Theoretical Background

Cognitive Load Theory

The basic assumption of the Cognitive Load Theory (CLT; van Merriënboer and Sweller 2005; Sweller 1988) is the limited capacity of working memory in terms of the amount of information that can be processed simultaneously and the time in which information is available for processing. The limitations of working memory cause a cognitive load on the learner in learning situations. We follow the current doctrine (e.g., Sweller et al 2019) and apply a two-factorial model for an intervention-induced cognitive load (see also Leppink et al 2013) so that the load is composed of intrinsic load (ICL) and extraneous or learning-irrelevant load (ECL). ICL refers to the complexity of information that the learner has to process during the learning process and is therefore determined by the learning task and the prior knowledge of the learner regarding the learning content. ECL refers to learning irrelevant cognitive processes, which occupy the working memory but do not lead to a relevant learning gain and can be influenced by the design of the learning procedure, such as how the information is presented to the learner. Principles can be derived from the CLT to enable instructors to design instructions that are conducive to learning (Sweller et al 2019). One fundamental principle is to keep the learning-irrelevant ECL as low as possible during the learning process (Leppink 2017; Leppink and van den Heuvel 2015; Sweller et al 1998). A negative learning effect, which results in an increase of ECL and has been empirically proven in numerous studies (see meta-analysis of Ginns 2006), is the split-attention effect. This effect postulates the necessity of cognitive integration

processes for the spatial and/or temporal separation of corresponding information sources, which leads to an increase in ECL and should accordingly be avoided in learning environments by physically integrating the information sources.

Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning (CTML; Mayer 1999, 2005; Moreno and Mayer 1999) is based on the CLT principle of a limited working memory, but, extending the CLT model, it meets three basic assumptions about information processing in working memory. First, the CTML postulates two separate channels of working memory in which auditory-verbal and visual-imagery information is processed (Dual-Channel Assumption). Second, the CTML assumes that the capacity of each channel is limited in terms of the amount of information that can be processed simultaneously (Limited Capacity Assumption). The third assumption is that the active engagement of the learner with the learning object itself is a necessary condition for the formation of a coherent mental model (Active Processing Assumption). From this theoretical approach, principles for the design of multimedia learning environments that promote learning can be derived (e.g., Mayer and Moreno 2003). In the following paragraphs, design principles are presented that can be fulfilled by the use of tablet-supported video analysis in teaching-learning situations in order to positively influence the learning process. The first of these is the contiguity principle, which is based on the avoidance of the split-attention effect and additionally distinguishes between the spatial and temporal separation of information.

Spatial Contiguity Principle

In accordance with this principle, learning environments should be designed in such a way that corresponding information is not presented to the learner spatially separated from each other. In this way, visual search processes that would bind cognitive resources of the learner without contributing to an increase in learning are avoided.

Temporal Contiguity Principle

This principle is based on avoiding the temporal separation of corresponding information sources. If this is not taken into account, the learner has to maintain a mental representation of the recorded information in the working memory in order to integrate it with the information that follows in time. This, in turn, ties up working memory resources without contributing to an increase in learning.

Another design principle of CTML is the segmentation or interactivity principle, which refers to dynamic visualizations in a multimedia learning environment.

Segmentation or Interactivity Principle

According to this principle, a learning-promoting effect is achieved by the fact that learners are not presented with information in a continuous unit but instead in discrete segments that they can call up one after the other (as required). This segmentation of the learning content avoids a cognitive overload of the learner, which could otherwise occur when the learner takes in too much information in too short a time. In this context, the term "interactivity principle" is also used by some authors (e.g., Robinson 2004). Interactivity can be understood as a possibility for the learner to control the presentation of the (multimedia) learning content. For example, the interactivity of a learning environment enables the learner to adapt the sequence and display duration of the information to be processed according to his own cognitive abilities.

Learning with Multiple Representations

For scientific learning, multiple external representations (MERs) play a beneficial role, which is well documented for the natural sciences (Tytler et al 2013) and for physics (Treagust et al 2017). It is especially important for conceptual understanding (Verschaffel et al 2010) and is discussed as a necessary condition for in-depth understanding (diSessa 2004). Ainsworth (2006, 2008) created a conceptual framework that provides an overview of the prerequisites for the effective use of MERs in teaching-learning situations and the unique benefits for learning complex or new scientific content. According to her, in the design, functions, and tasks (DeFT) taxonomy, learning with MERs means two or more external representations (e.g., diagrams, formulas, and data tables) are used simultaneously. Specifically, there are three key functions that MERs can fulfill (even simultaneously) to support the learning process. According to the first function, MERs can complement one another either by providing complementary information or by allowing for complementary approaches to process information. The presentation of MERs with complementary information content may be advantageous if the presentation of all relevant information in a single form of representation would lead to cognitive overload of the learner. However, even if different representations contain the same information, they can support the learning process by allowing learners to select the most appropriate form of representation to accomplish the given learning task in the particular learning situation. The second function is that simultaneously presented rep-

resentations can constrain one another's interpretation in two ways: The more familiar representation can constrain the interpretation of the less-familiar one, or the inherent properties of one representation can trigger the usage of the other representation, which is considered helpful for the learning process. According to the third function, the construction of a deeper understanding is fostered if learners integrate information from different forms of representation to gain insights that could not have been obtained with just one form of representation. Although MERs demonstrably have the potential to support learning processes, their use in learning situations is also associated with learner demands that can increase learners' cognitive load and even negatively affect the learning process (de Jong et al 1998). Indeed, many studies point toward student difficulties with MERs (e.g., Ainsworth 2006; Nieminen et al 2010). Consequently, the cognitive load of the learning environment must be considered and managed carefully. In this context, technology support can help reduce cognitive load when learning with MERs, and it can therefore facilitate the learning-promoting effect of MERs (e.g., Horz et al 2009).

The Role of Emotions in Learning Processes

The influence of motivational and emotional processes, especially on technology-supported learning, has been insufficiently considered in research for a long time. It was Moreno who first systematically incorporated these factors into the theoretical model for multimedia learning by, together with Mayer, extending CTML to include motivational, affective, and metacognitive factors that influence information processing. In the resulting Cognitive-Affective Theory of Learning with Media (CATLM; Moreno 2005; Moreno and Mayer 2007), "motivational factors mediate learning by increasing or decreasing cognitive engagement" (p. 151 Moreno 2006). Metacognitive factors, on the other hand, have an indirect influence on learning success by moderating the regulation of cognitive processing, motivation, and emotions. Following this line of argument, Plass and Kaplan (2016) conclude from the inherent connection between emotion and cognition that all information processing in a learning environment is both emotional and cognitive. For the control of learning processes, it is therefore essential to consider the constant dynamic interplay between cognition and emotion during the individual learning steps, especially for digitally supported learning situations, since "digital learning environments offer many more ways of influencing learners' emotions" (p. 137 Plass and Kaplan 2016). In order to describe the effect of emotions in learning processes theoretically, Plass and Kaplan extended the CTML by a mutual relationship between emotion and cognition by including emotions as a separate processing channel into the ICALM model.

The Integrated Cognitive Affective Model of Learning

In the Integrated Cognitive Affective Model of Learning (ICALM), the multimedia learning environment induces affective reactions. Selection processes taking place in the working memory and organizational processes of visual or auditory information are influenced by affective variables such as situational interest or learning motivation and vice versa. Mental representations are thus linked to the corresponding affective variables and are integrated into long-term memory as emotional schemata. Plass and Kaplan thus assume that emotional self-regulation mechanisms can occupy cognitive resources of the working memory in learning situations and thus result in additional cognitive load. According to Plass and Kalyuga (2019), one possibility to describe the influence of emotions on the cognitive load in this model is to understand the processing of emotions during the learning process as additional ECL. In principle, two mechanisms are distinguished for this (Oaksford et al 1996). On the one hand, emotions can be caused in learning environments whose regulation binds cognitive resources of the working memory but does not lead to an increase in learning (Pekrun 2000). An example illustrating this is the effect of stress generated by pressure to perform in a learning situation. In such a situation, the working memory also processes thoughts about a possible failure so that less cognitive resources are available for an active knowledge construction (Beilock et al 2004). This effect, which can occur with both positive and negative emotions, has been empirically proven in numerous studies, especially for negative emotions (e.g., Brand et al 2007; D'Mello and Graesser 2012). The second mechanism describes the emotionally induced processing of information that has no relevant meaning for the achievement of the learning goal. In this case, the processing puts an additional load on the working memory but does not lead to an increase in learning and thus increases ECL. This occurs, for example, if information irrelevant to learning appears more interesting to the learner by adding certain details, but this leads to a lower learning performance (seductive details effect, Harp and Mayer 1998). The processing of these interesting details results in an additional cognitive load that is irrelevant to learning and whose negative effect on learning performance overcompensates for the effect of increased attention. In both of the emotionally induced mechanisms of action described above, resources of the working memory are thus occupied, which are not available to the learner for an active knowledge construction in a learning situation. Recalling the fundamental demand of CLT to keep ECL in

multimedia learning environments as low as possible, emotional processes should be considered when designing such learning environments, and the effect on the emotional state of the learners should be examined.

Learning Effectiveness of Tablet PC-supported Video Analysis

In the following paragraphs, the learning effectiveness of the digital tool is founded on the fundamental learning theories described above. In particular, which elementary functions of the video analysis application can contribute to a learning-promoting use of MERs in learning environments will be shown.

*Fulfillment of the Key Functions of the DeFT Framework* The video analysis application automatically provides learners with multiple forms of representation that include complementary information regarding the investigated motion, the real and stroboscopic image, and the associated

motion diagrams (the first key function). The learners in this study are much more familiar with the diagram representation form than with the stroboscopic image representation form because of the experienced teaching (in accordance with their curriculum). The interpretation of the less-familiar stroboscopic image could thus be triggered by the automatic display together with the familiar motion diagram (the second key function). Also, the automatic multicoding of the motion process makes it easier for the learner to integrate the information from the different MERs into a coherent mental model of the motion process, which would not be possible if only one form of representation were presented (the third key function).

Furthermore, the automatic display of MERs by means of the video analysis application fulfills the basic design principles of the CTML described above.

Fulfillment of the Principle of Contiguity The video analysis application displays real motion sequence and stroboscopic image time-synchronously as well as one-dimensional time-position and time-velocity diagrams in

Fig. 1 Screenshots of the video analysis application Viana; real image and superimposed stroboscopic image (*top*), x(t) diagram (*middle*), and v_x(t) diagram (*bottom*)
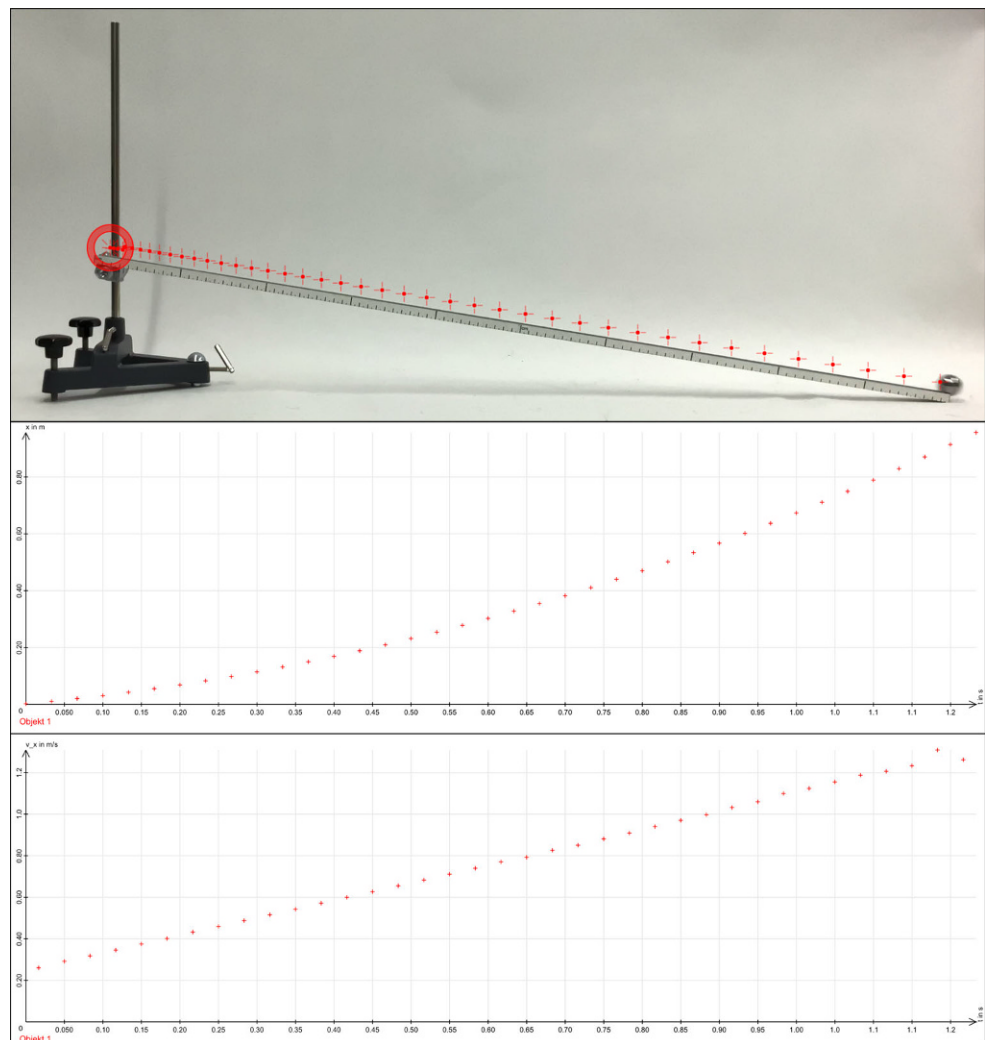
one combined image (see Fig. 1). This avoids a spatial separation of corresponding information and thus fulfills the spatial contiguity principle. The learner can also switch between these images by wiping the screen of the tablet PC. This intuitive hand gesture, with which learners are also familiar due to the omnipresence of digital media today, allows different forms of representation to be called up quasi-simultaneously, which means by a gesture of the hand without noticeable time delay. Corresponding information is thus presented to the learner without temporal separation, and the temporal contiguity principle is fulfilled.

Fulfillment of the Interactivity Principle During the video analysis, the students themselves can control the transition between the individual segments (forms of representation) of the data evaluation (real motion sequence, motion diagrams, and the stroboscopic image). For example, to improve the understanding of the motion diagrams, the students can again call up the stroboscopic image or vice versa. The use of the video analysis application fulfills the interactivity principle by allowing this self-control of the learning process. The interactivity of the application also promotes the active engagement of the learners with the subject of learning itself. For example, the origin and spatial orientation of the coordinate system can be manipulated by the learner at any point in the learning process, and the effects of this variation on the motion diagrams can be observed quasi-simultaneously. According to Ainsworth (2006), this dynamic linking of different forms of representation can contribute to a reduction of extraneous cognitive load on learners.

In summary, the visualization features of the video analysis application regarding MERs meet the design principles for multimedia learning environments and thereby reduce ECL (especially via the avoidance of the split-attention effect) in multirepresentational learning environments; in turn, this fosters the effective use of MERs in the study's learning scenarios.

Research Hypotheses and Research Questions

This work focuses on the learning effects of the digital tool in experiment-based learning. The corresponding results of the preliminary study prove a greater learning-promoting effect of tablet-supported video analysis compared to traditional teaching sequences based on measurement and analysis tools established in school education (stopwatch, tape measure, and graphing calculator). In this way, positive effects for the promotion of conceptual understanding could be demonstrated, the reproducibility of which was examined in the replication study.

Research Hypothesis 1 Experiment-based teaching sequences, based on tablet PC-supported video analysis, lead to an increased learning gain in terms of conceptual understanding compared to traditional teaching sequences. Positive effects of the preliminary study should therefore be reproduced.

The theoretical foundation of the learning effectiveness of tablet-supported video analysis as presented in this paper is based on the reduction of extraneous load through the multirepresentational visualization possibilities of the video analysis application used. A further objective of the replication study is to prove the reduction of the extraneous load and to empirically verify the theoretical foundation of the learning effectiveness.

Research Hypothesis 2 Experiment-based teaching sequences, based on tablet-supported video analysis, lead to a significant reduction of the extraneous load compared to traditional teaching sequences.

Research Hypothesis 3 The theoretically founded causal relationship between reduction of extraneous cognitive load and increased conceptual understanding should be empirically verifiable.

Hillmayr et al (2020) demonstrated in a meta-study that the use of digital media in regular school lessons across all examined subjects (biology, chemistry, physics, and mathematics) leads to an increase in motivation and a more positive attitude toward the subject involved. Sung et al (2016) could also show in a meta-study that the use of mobile devices in the educational context has a positive effect on affective variables (e.g., motivation, engagement, attitude, satisfaction, and preference). Technological support is therefore inferred to have a positive influence on the emotional state of learners during the learning process. With the results of the preliminary study, however, it is not possible to make a statement about the influence of technological support on emotional variables. Accordingly, the replication study also aimed to investigate the influence of tablet-supported video analysis on emotional variables.

Research Question 1 Do teaching sequences based on tablet-supported video analysis have a positive influence on the emotional state of the learners compared to traditional teaching sequences?

According to the ICALM model, there is a direct relationship between emotions and cognitive load. Another objective of the replication study was therefore to empirically test this theoretically founded relationship.

Research Question 2 Can a causal relationship between intervention-induced emotions and extraneous cognitive load be empirically verified?

Study Design

The study design of the preliminary study was adopted for the replication study. Accordingly, a cluster-randomized controlled trial involving high school physics courses was conducted. As a consequence, whole courses were assigned as treatment group (TG) or control group (CG). As with the preliminary study, the replication study also covered two subject areas, the uniform motion (UM) and the accelerated motion (AM). Thus, the two teaching sequences on the two topics were sequential, starting with the teaching sequence on the uniform motion.

Experimental Manipulation

The students in the TG first recorded the motion process with a tablet PC and then analyzed and visualized the measurement data with the video analysis application. The students in the CG, on the other hand, used experimental tools commonly used in traditional school teaching. They acquired the measured values with a stopwatch and tape measure, entered them into a graphing calculator (TI-84 Plus, Texas Instruments), and used the calculator to analyze and visualize the measured values. Great importance was attached to ensuring a fair comparison between the groups. Thus, for both groups, the experimental setups, the learning time, the learning content, and the social forms of learning were identical.

Teaching Sequences

The teaching sequences for both subject areas had the same structure, which is shown in Fig. 2. In the first lesson of the sequence, a pre test was carried out in both the CG and the TG in order to empirically assess the prior knowledge of the students regarding the respective subject areas. The

subsequent learning phase is divided into an experiment-based learning phase of four 45-minute lessons and a consolidating exercise phase of two 45-minute lessons. Immediately following the respective learning phase, testing was repeated. In order to achieve the highest possible degree of comparability of the learning gain in the individual learning phases, the achievement tests contained identical items at each test time, but in a different order.

Introductory Lesson

In the first lesson of the intervention, the teachers first explained the organizational conditions and the procedure of the teaching sequences to the learners. Subsequently, the students of the TG received standardized instruction in video motion analysis with the tablet. The briefing included an explanation of the measurement method as well as the functionality and handling of the video analysis application. Additionally, the students were given the opportunity to familiarize themselves with the application used by analyzing sample videos. The sample videos are already included in the library of the video analysis application and have no contextual connection to the topics of the study. The teacher provided a short description of the functional range of the application as operating instructions. The students of the CG received standardized instruction in data analysis with the graphing calculator and were familiarized with the functionality by analyzing a given data set. A short description of the data processing with the graphing calculator was handed out by the teacher.

Experiment-based Learning Phase

In this learning phase, students in both groups experimented collaboratively in small groups. The group sizes were set to a minimum of two and a maximum of three. The students
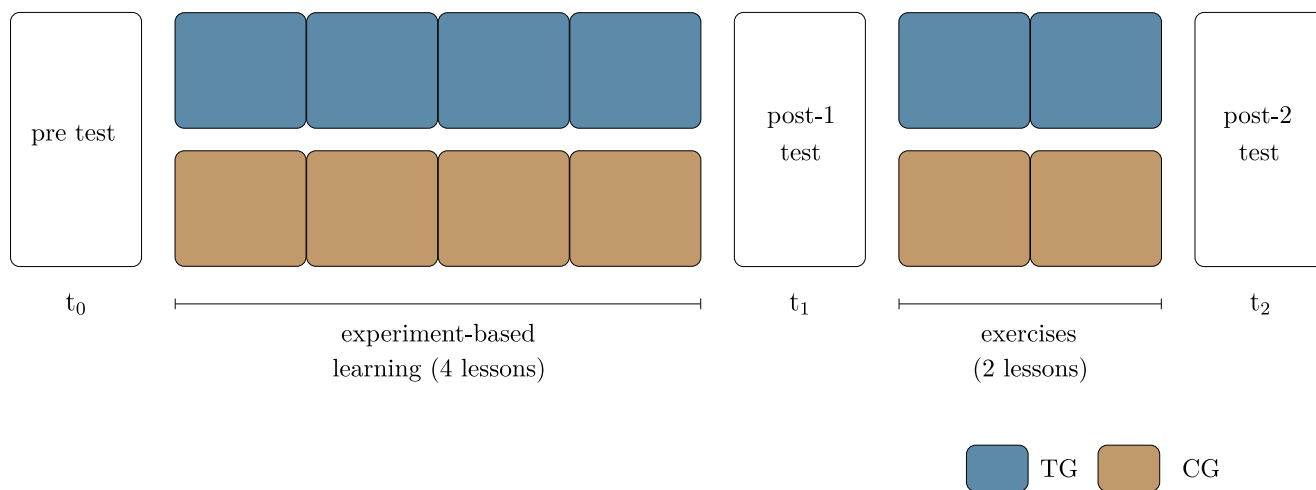

Fig. 2 Structure of a teaching sequence (each lesson lasts 45 minutes)

were given learning tasks whose central learning object was the experiment. By means of a standardized experiment instruction, the students were guided to an autonomous execution and evaluation of the measured data, and thus, the influence of the teacher was reduced as much as possible. The teachers were also instructed to take a passive role during the experiment to accompany the learning process and not to actively intervene. This was intended to ensure that the learning gain resulted from the independent processing of the learning tasks and not from interaction with the teacher. The experiments were developed under the following two premises. Firstly, the measured values should be equally well measurable by video analysis and by conventional experimental tools in the same period of time so that a truly fair comparison between the two groups is guaranteed. Secondly, experiments should be carried out with everyday, inexpensive materials. This should ensure that these experiments can be performed in regular lessons without great additional financial outlay in order to promote the transfer to school practice. Thus, only an aluminum rail, a steel ball, and a piece of rubber were needed for the experiments. It should be noted that the friction between the steel ball and the rail during the motion was so low that it could be neglected for the physical description of the motion of the rolling ball.

Uniform Motion The experimental phase included a total of two experiments, for whose setup, execution, and evaluation the students had two lessons each. The students of the TG carried out the measurement of the time-dependent position data of the rolling ball as well as the determination and visualization of the velocity using video analysis. The students of the CG measured the position data with a stopwatch and tape measure, entered the measured values into the graphing calculator, and determined and visualized the velocity graphically with the calculator.

Experiment 1 The first experiment had as its central learning object the conceptual understanding of velocity regarding a uniform motion in one direction. The students let the steel ball roll over the aluminum rail lying flat on the table, once fast and once slow.

Experiment 2 The second experiment focused on the conceptual understanding of velocity in a uniform back and forth motion. The students rolled the steel ball once in one direction, stopped it, and then rolled the ball back again in the other direction. The educational background is that in physics, velocity is a vectorial quantity that can have a positive or negative sign, depending on the reference system and the direction of the motion, which can be associated with learning difficulties for students (McDermott et al 1987).

Accelerated Motion As before, the students carried out two experiments in two lessons each. For the measurement of the time-dependent position data of the rolling ball, the students of the TG used video analysis, and the students of the CG used a stopwatch and a tape measure. Velocity and acceleration were also determined and visualized by video analysis in the TG, while the CG used the graphing calculator for this purpose.

Experiment 1 The learning content of the first experiment was the conceptual understanding of acceleration regarding a motion in one direction with constant acceleration. The students fixed a piece of rubber under one end of the rail to obtain a small inclination. They then let the ball roll down from the raised end, keeping the inclination constant.

Experiment 2 The second experiment concentrated on the laws of motion in uniformly decelerated motion and their distinction from uniformly accelerated motion. Halloun (2006) emphasizes in this context the importance of the learning objective for students to link these types of motion with the physical quantity of acceleration. For this purpose, the students rolled the ball up the rail and determined the time-dependent position of the ball both when rolling up and down. In this way, the vectorial character of the physical quantity acceleration should be cognitively grasped, especially the resulting distinction into an accelerating or decelerating effect on moving bodies.

Exercise Phase In order to stay as close as possible to the course of a real teaching sequence with the study, the students worked on exercises after the experiment, which aimed at consolidating and deepening the declarative and procedural knowledge acquired in the experiment-based phase. Exercise booklets were handed out for this purpose, in which the students also entered their solutions to the tasks. The exercises were carried out collaboratively in small groups of two to three students. While the CG worked on traditional, paper-based tasks, the TG analyzed ready-made videos of the corresponding motion type, but otherwise, the exercises were designed comparably for both groups. In order to enable the teachers to take a passive role in this phase as well, the students were given the opportunity to compare their solutions with sample solutions after completing the tasks. The exercises were designed similarly for both groups, so that a fair comparison between the two groups was also possible in this phase.

Instruction Material

At the beginning of the intervention, the instructions for carrying out the experiments and the associated learning tasks were handed out as material in a protocol booklet, into which the students also entered their results. When developing the instructional materials in cooperation with teachers with many years of professional experience, great importance was attached to the comparability of the materials for both groups in terms of learning content, forms of representation used, volume, and level of difficulty. In particular, the instruction material was designed in such a way

Table 2 Sociodemographic composition of the sample population

	CG	TG
Sample size	136	150
Average age	15.7	15.6
Female (in %)	31.9	34.0
Physics advanced course (in %)	40.0	39.0
Mathematics advanced course (in %)	37.7	38.9

Table 3 Sub-concepts of the achievement tests

Sub-concept	Content	No. of items
Uniform motion		
G1	Velocity as alteration rate	3
G2	Velocity as vectorial quantity	3
G3	Reference system	3
Accelerated motion		
B1	Acceleration as alteration rate	3
B2	Acceleration as vectorial quantity	3
B3	Reference system	5

that the learners in both groups were able to carry out and evaluate the experiments in small groups as independently as possible within the scheduled teaching time and to master the exercises.

Methodology

Sample

The data was collected in 18 courses from 11 secondary schools in different states in Germany between 2017 and 2018. In total, 294 students participated in both test times. Sociodemographic data was evaluated for 286 students, of whom 94 are female and 191 are male (one did not complete the question about gender), with an average age of 15.6 ($SD = 0.72$). The sociodemographic composition separated according to TG and CG is shown in Table 2.

Instruments

All instruments used were subjected to confirmatory factor analysis to empirically confirm the intended factor structure after checking the necessary prerequisites (KMO, Bartlett's Test of Sphericity). The response patterns of the combined sample at the respective post time point were analyzed with the software R and the R-package lavaan (version 0.6-3). The corresponding results and the questionnaires used can be found in the Electronic Supplementary Material.

Conceptual Understanding

In order to detect the effects in learning achievement regarding conceptual understanding, the achievement test in multiple choice design already evaluated in the preliminary study was used again. Thus, the structuring into three sub-concepts (G1, G2, and G3 or B1, B2, and B3) per subject area was adopted (see Table 3).

Cognitive Load

The cognitive load induced by the intervention was measured at the post time point of the respective learning phase with a 10 item questionnaire developed and vali-

dated by Leppink et al (2013). With this instrument, the cognitive load can be measured differentiated according to intrinsic (ICL), extraneous (ECL), and germane cognitive load (Hadie and Yusoff 2016; Zukić et al 2016). Although a two-factorial model of cognitive load was used in this work, the test instrument as a whole was used to ensure the validity of the measurement. For further analysis, however, only response patterns of items concerning ICL and ECL were used. The items were first translated into German and then adapted to the context and the respective learning phase (experiment-based and exercise phase). Finally, the number of possible options per item was reduced from 11 to six in order to maintain consistency with the other questionnaires used.

Emotions

The emotional state of the learners was captured with a five-item questionnaire immediately after the respective learning phase, asking about the learners' subjective assessment of positive-activating (pleasure, satisfaction) and negative-deactivating emotions (boredom, frustration, and uncertainty) during the learning process. This two-dimensional classification of achievement emotions is based on the work of Pekrun (2014). Items were taken from the Achievement Emotions Questionnaire developed and validated (Pekrun et al 2002, 2011) and later translated into German.

Teacher Behavior

The behavior of the teacher during the intervention was assessed at the post time point of the respective learning phase using a four-item scale already applied in the preliminary study. The students were asked about the teacher's commitment, willingness to support, and motivating effect. By means of a confirmatory factor analysis, a two-factor structure was identified, evoking a separation into two sub-scales: willingness to support (WS) and commitment (CM).

Analysis Methods

Multilevel Regression Analysis

Since it cannot be excluded that the learning process of the students is influenced by the social group to which they belong, the hierarchical data structure with two levels (student and course level) was taken into account when choosing the method of analysis. For this purpose, two-level regression analyses, according to Hox (2010), has been carried out to detect group effects with 18 clusters on level two. For a multilevel analytical procedure, however, this number must be regarded as small. In this case, Hox and McNeish (2020) recommend the use of specially developed procedures to avoid bias in the estimates of standard errors and thus improve the accuracy of parameter estimation for small sample sizes. Consequently, the most common procedure for parameter estimation in multilevel regression models, the Full Information Maximum Likelihood Method (FIML), should be replaced by the Restricted Maximum Likelihood Method (REML). In this method, the variance components are first estimated without taking the fixed effects into account, which results in a more accurate estimation of the variance components. Then, the fixed effects are estimated as a function of these variance components. Furthermore, a correction should be made to the significance tests for fixed effects. A common procedure for this is the so-called Kenward-Roger correction (Kenward and Roger 1997). In this correction procedure, the bias due to low sample sizes is first estimated at higher levels, and then the standard errors are corrected for this bias. In addition, the degrees of freedom are approximated based on the parameter estimates for the model under consideration. If both methods are combined, simulation studies show a significant reduction of bias due to small sample sizes at higher levels (e.g., Luke 2017). In this case, Hox and McNeish specify based on the results of simulation studies (McNeish and Stapleton 2016) a minimum number of five to eight clusters in cross-sectional designs "for estimates to be stable and trustworthy" (p. 218). The multilevel regression analyses for this contribution were carried out using the software R and the R-package lme4 (version 1.1.21). Another advantage of this analysis method is that the influence of the aggregation in courses can be quantified and thus assessed by the proportion of variance explained at level two. The associated measure is the so-called Intraclass Correlation Coefficient (ICC), for whose calculation the R-package sjstats (version 0.17.7) was additionally used.

Structural Equation Modeling

To empirically verify the hypothetical causal relationships between emotion, cognitive load, and learning achievement, and to investigate the influence of teacher behavior, the method of structural equation modeling (see e.g., Kline 2011; Jöreskog 1978) was applied. All necessary analysis steps were performed with the software R and the R-package lavaan (version 0.6-3). Structural equation modeling is a statistical technique that requires a large sample size. A recommendation frequently given in the academic literature is a minimum sample size of $N = 200$ (e.g., Kline 2011; Barrett 2007), but based on simulation studies, other authors estimate a sample size of $N = 100–150$ as acceptable (e.g., Anderson and Gerbing 1984; Muthén and Muthén 2002). Since the given data was ordinally scaled and not multivariate normally distributed, the model parameters were estimated using the Diagonal Weighted Least Squares (DWLS) method. Since the sample size achieved in this study can be considered small for structural equation analysis, the DWLS method was used in its robust variant, which is more suitable for small samples. In this variant, the DWLS procedure is used to estimate the model parameters, but the standard errors are corrected, and a mean and variance-adjusted test statistic is used.

Results

Since this work focuses on the learning effect of the digital tool in the experiment-based learning processes, the results for the experimental phase are shown and discussed here. However, the results for the exercise phase can be found in the Electronic Supplementary Material.

Preliminary Analyses

Group Differences Before the Intervention Began

In order to identify significant differences that existed between the groups before the intervention began, the Mann-Whitney U test was conducted for course selection, preliminary grades, and prior knowledge (two-sided, significance level 5%). The results are shown in Table 4. There are

Table 4 Test for significant group differences

| Variable	TG *M(SD)*	CG *M(SD)*	U test *p*
Course choice			
Physics	1.61 (0.49)	1.60 (0.49)	0.900
Mathematics	1.61 (0.49)	1.62 (0.49)	0.800
Preliminary grade			
Physics	2.26 (1.20)	2.63 (1.41)	0.040
Mathematics	2.54 (1.25)	3.01 (1.37)	0.004
German	3.05 (1.09)	3.25 (1.15)	0.100
Conceptual understanding			
Pre test	4.66 (2.09)	3.92 (1.99)	0.004

significant group differences for the preliminary grades in physics ($p = 0.040$) and mathematics ($p = 0.004$) as well as prior knowledge ($p = 0.004$). In order to control these group differences statistically, a sample balanced in these covariates was generated with the method of Propensity Score Matching (PSM; e.g., Rosenbaum and Rubin 1983; Guo and Fraser 2010) prior to the following comparative analysis.

Propensity Score Matching

PSM allows causal statements on intervention effects to be made, even in empirical studies in which complete randomization was not sufficiently successful (Fan and L. Nowell 2011). For all potential confounding variables, the so-called propensity score (PS) is determined for each individual respondent using a logistic regression model, and thus, each respondent in one group is assigned one or more respondents in the other group with the same or very similar PS values. The data set of this paired population can then be analyzed using conventional statistical methods. After the matching process, the population was balanced in all covariates (see Table 5), but the sample size was reduced from $N = 286$ to $N = 262$. However, this was still sufficient for the statistical analysis procedures intended in this work, so that all further analyses will be based on the data set of the paired sample.

Table 5 Test for significant group differences after PSM

Variable	TG M(SD)	CG M(SD)	U test p
Course choice			
Physics	1.66 (0.47)	1.64 (0.48)	0.700
Mathematics	1.66 (0.48)	1.65 (0.49)	0.900
Preliminary grade			
Physics	2.34 (1.20)	2.61 (1.40)	0.157
Mathematics	2.71 (1.22)	3.00 (1.37)	0.101
German	3.05 (1.13)	3.24 (1.16)	0.157
Conceptual understanding			
Pre test	4.31 (1.94)	3.95 (2.01)	0.198
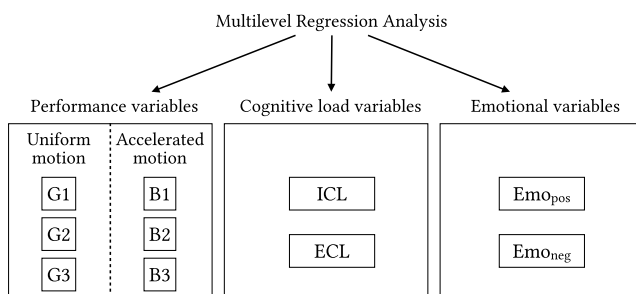
Fig. 3 Schematic illustration of the analysis process

Multilevel Regression Analysis

For the comparative analyses between the treatment group and the control group, the cognitive performance variables, the cognitive load variables, and the emotional variables regarding the respective subject areas were subjected to multilevel regression analysis as shown in Fig. 3.

Conceptual Understanding

In a first step of the multilevel regression analysis, the ICC value was determined. For both topics, it could be found that only a small proportion of the variance is localized at the course level ($ICC_{G1} = 9.4\%$, $ICC_{G2} = 7.2\%$, $ICC_{G3} = 12.1\%$; $ICC_{B1} = 1.8\%$, $ICC_{B2} = 3.8\%$, $ICC_{B3} = 9.4\%$). Fig. 4 gives an overview of the group-dependent mean values and standard errors of the relative test scores, differentiated by the sub-concepts. In order to reveal significant differences in learning gain between the two groups, the next step was to examine whether there was an interaction effect between the time of testing and group affiliation regarding the different sub-concepts. The results are shown in Table 6. Significant effects in favor of the TG were demonstrated for two sub-concepts ($G3$: $p = 0.002$, $\eta_2 = 0.020$, $1 - \beta = 0.879$; $B2$: $p < 10^{-3}$, $\eta^2 = 0.053$, $1 - \beta = 0.974$) in favor of the CG for one sub-concept ($B1$: $p = 0.012$, $\eta^2 = 0.023$, $1 - \beta = 0.654$). For comparison, the effects of the preliminary study are shown in Table 7.

Cognitive Load

For ICL, as for the achievement variables, only a small part of the variance is localized above the student level [$ICC_{ICL} = 3.6\%$ (UM)/9.8% (AM)]. This is also true for ECL with respect to the accelerated motion ($ICC_{ECL} = 7.4\%$), but regarding the uniform motion, a comparatively higher proportion of the variance is explained on level two ($ICC_{ECL} = 24.9\%$). The examination for group differences (see Table 8) yielded a significantly lower extraneous load for the TG for both subject areas (UM: $p = 0.013$, $\eta^2 = 0.034$, $1 - \beta = 0.831$; AM: $p = 0.003$, $\eta^2 = 0.090$, $1 - \beta = 0.979$). In addition, however, the intrinsic load regarding the accelerated motion was significantly lower for the TG ($p = 0.029$, $\eta^2 = 0.041$, $1 - \beta = 0.748$).

Emotions

The proportion of variance explained at level two is estimated to be low for positive emotions for both subject areas, but higher than for cognitive load and achievement variables [$ICC_{Emo_{pos}} = 10.1\%$ (UM)/15.1% (AM)]. While for the subject area of accelerated motion this also applies to negative emotions ($ICC_{Emo_{neg}} = 14.3\%$), the proportion of the ex-
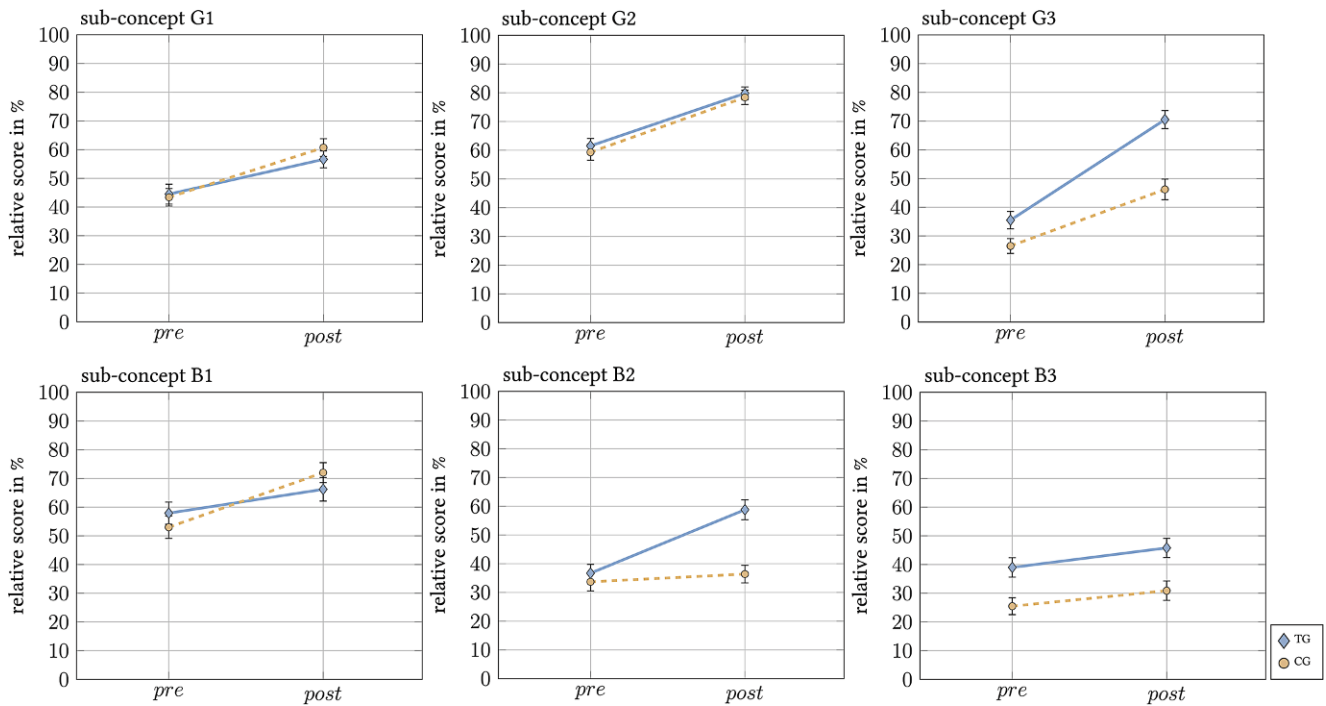
Fig. 4 Scores of the achievement tests differentiated by sub-concepts and group

Table 6 Effects of the replication study on experiment-based learning

Sub-concept	ICC	F	p	η^2	$1 - \beta$
Uniform motion ($N = 244$)					
G1	0.094	1.429	0.233	–	–
G2	0.072	.040	0.842	–	–
G3	0.121	9.805	0.002	0.020	0.879
Accelerated motion ($N = 168$)					
B1	0.018	5.563	0.012	−0.023*	0.654
B2	0.038	15.318	$< 10^{-3}$	0.053	0.974
B3	0.094	0.074	0.787	–	–

* The negative sign indicates an effect in favor of the control group

Table 7 Effects of the preliminary study on experiment-based learning

Sub-concept	ICC	F	p	η^2	$1 - \beta$
Uniform motion ($N = 109$)					
G2	–	–	0.026	0.036	0.51
G3	–	–	$< 10^{-3}$	0.093	0.91
Accelerated motion ($N = 70$)					
B1	–	–	0.022	0.050	0.46
B2	–	–	0.001	0.138	0.90

Table 8 Test for group differences in cognitive load

Cognitive load	ICC	F	p	η^2	$1 - \beta$
Uniform motion ($N = 233$)					
ICL	0.036	.545	0.472	–	–
ECL	0.249	7.920	0.013	0.034	0.831
Accelerated motion ($N = 164$)					
ICL	0.098	6.564	0.029	0.041	0.748
ECL	0.074	15.806	0.003	0.090	0.979

Table 9 Test for group differences in emotional states

Emotions	ICC	F	p	η^2	$1 - \beta$
Uniform motion ($N = 228$)					
Emo$_{pos}$	0.101	1.814	0.199	–	–
Emo$_{neg}$	0.243	8.737	0.010	0.046	0.927
Accelerated motion ($N = 160$)					
Emo$_{pos}$	0.151	2.439	0.150	–	–
Emo$_{neg}$	0.143	9.303	0.013	0.057	0.878

plained variance for the subject area of uniform motion is comparatively higher ($ICC_{Emo_{neg}} = 24.3\%$), as it is for ECL. A group comparison (see Table 9) showed that the students in the TG developed significantly lower levels of negative-deactivating emotions in the teaching sequences of both subject areas (UM: $p = 0.010$, $\eta^2 = 0.046$, $1 - \beta = 0.927$; AM: $p = 0.013$, $\eta^2 = 0.057$, $1 - \beta = 0.878$).

Structural Equation Modeling

Measurement Model

Based on the confirmatory factor analyses of the response patterns of the instruments used, a measurement model was

generated to operationalize the latent variables positive-activating emotions (Emo$_{pos}$), negative-deactivating emotions (Emo$_{neg}$), intrinsic (ICL) and extraneous (ECL) cognitive load, and learning gain regarding the sub-concepts G3 and B2 as well as the teacher behavior variables willingness to support (WS) and commitment (CM).

Structural Model

The direct causal relationship between emotion, cognitive load, and learning achievement, which is based on the CLT and ICALM models, was transferred into a structural model. Teacher behavior was then integrated into the model as an additional influencing variable, as it cannot be excluded that this may influence the induced emotions as well as the extraneous load. Thus, the structural model is based on the following hypotheses:

- The higher the level of negative emotions, the higher the extraneous load.
- The higher the level of positive emotions, the lower the extraneous load.
- The higher the extraneous load, the lower the learning achievement.
- Teacher behavior influences induced emotions and extraneous load.
- The higher the teacher's commitment, the greater his or her willingness to support.

The resulting structural model is illustrated in Figs. 5 and 6 as a path diagram with color-coded significant paths for the two subject areas.
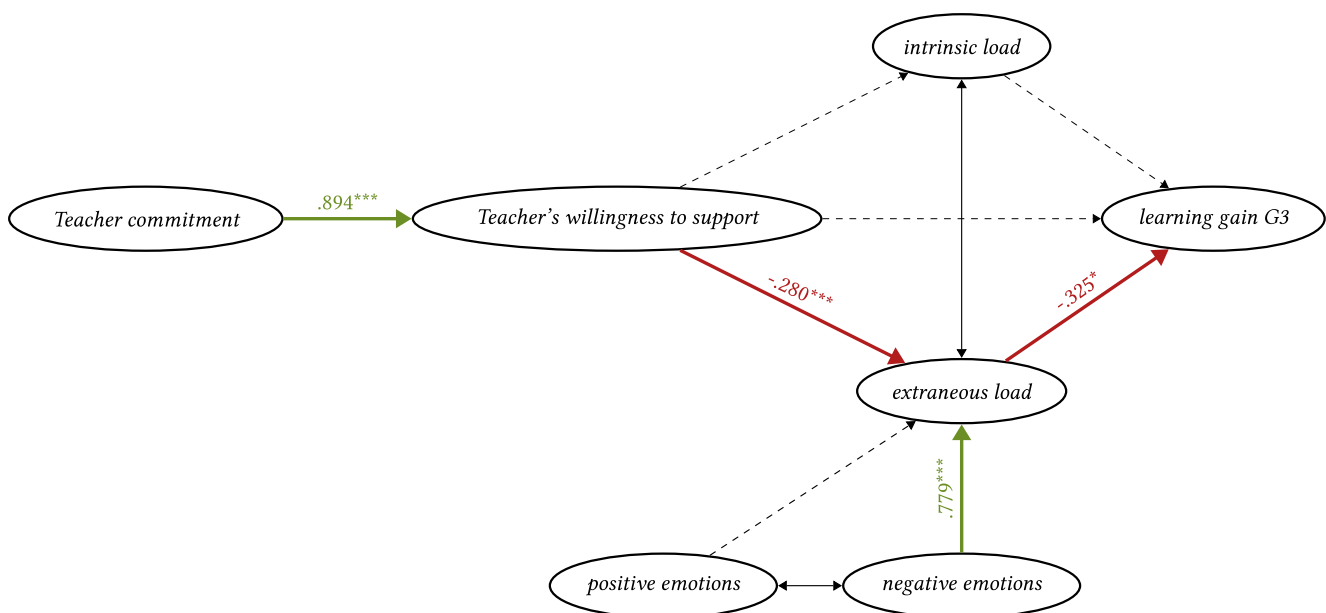
Uniform Motion

The resulting model with 47 free parameters fits the given data well, $p(\chi^2) = 0.044$, CFI $= 0.981$, TLI $= 0.976$, RMSEA $= 0.033$, SRMR $= 0.049$. The significant path coefficients are listed in Table 10.

Accelerated Motion

The resulting model with 47 free parameters fits the given data well, $p(\chi^2) = 0.146$, CFI $= 0.997$, TLI $= 0.996$, RMSEA $= 0.030$, SRMR $= 0.072$. The significant path coefficients are listed in Table 11.

Discussion

Replication of Effects of the Preliminary Study

For the subject area of uniform motion, the positive effect with respect to the sub-concept "reference system" (G3) was again demonstrated with good test power. Effect size and test power are lower in the replication study than in the preliminary study, which can possibly be attributed to the more advanced analysis method, since it also takes the hierarchical data structure into account. One possible explanation for this replicated positive effect is the interactivity of the video analysis application, which allows students to proactively manipulate the origin and spatial orientation of the coordinate system at any point in the learning process. By referring back to fundamental learning theories (CLT,

Fig. 5 Structural model with color-coded significant paths for the uniform motion
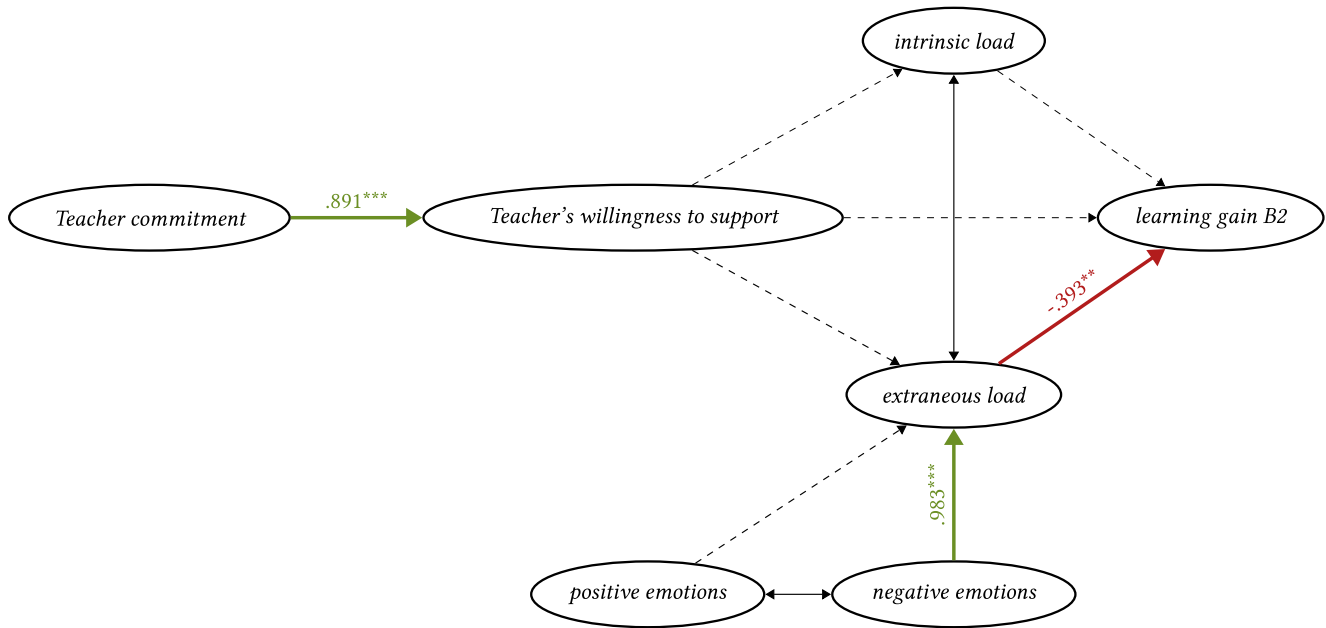
Fig. 6 Structural model with color-coded significant paths for the accelerated motion

Table 10 Significant path coefficients for uniform motion

Outcome	Predictor	b	se	z	p	β
G3	ECL	−0.324	0.136	−2.381	0.017	−0.325
ECL	Emo_{neg}	0.704	0.071	9.941	$< 10^{-3}$	0.779
ECL	WS	−0.235	0.065	−3.603	$< 10^{-3}$	−0.280
WS	CM	0.941	0.098	9.644	$< 10^{-3}$	0.894

$N = 226$; $b =$ unstandardized estimate, $\beta =$ standardized estimate

Table 11 Significant path coefficients for accelerated motion

Outcome	Predictor	b	se	z	p	β
B2	ECL	−0.417	0.156	−2.672	0.008	−0.393
ECL	Emo_{neg}	1.130	0.197	5.746	$< 10^{-3}$	0.983
WS	CM	.839	0.071	11.787	$< 10^{-3}$	0.891

$N = 157$; $b =$ unstandardized estimate, $\beta =$ standardized estimate

CTML), the active engagement with the learning object and thus the construction of knowledge is promoted. On the other hand, the dynamic linking of the coordinate system and the motion diagrams, according to Ainsworth (2006), contributes to the reduction of extraneous load, leading to an increase in learning achievement.

For the subject area of accelerated motion, the positive effect with respect to the sub-concept "acceleration as a vectorial quantity" (B2), which was demonstrated with high test power in the preliminary study, could be replicated with even higher test power. The effect size in the replication study was lower than in the preliminary study, which, again, could be a consequence of the analysis procedure. The interactivity of the video analysis application and the dynamic linking of coordinate system and motion diagrams

can also be used to explain this effect. To understand acceleration as a vectorial quantity, it is essential to distinguish a positive from a negative acceleration. The sign, however, depends on the selected coordinate system, so that the above mentioned advantages of active manipulation and dynamic linking of the coordinate system could also have a learning effect on the understanding of this sub-concept.

In contrast, effects that could only be detected with a low test power in the preliminary study were no longer statistically detectable for the replication study. In particular, even a negative effect was identified in the replication study with regard to sub-concept "acceleration as alteration rate" (B1), whereas a positive effect was found in the preliminary study. This shows the necessity of taking test powers into account to assess the statistical significance, and thus, the

validity of study results, especially in empirical educational research (Shadish et al 2002).

It is noticeable that the students of the TG showed a significantly higher learning gain regarding the sub-concept "vectorial character of the physical quantity" in the subject area accelerated motion but not in the subject area uniform motion. One possible explanation lies in the students' greater prior knowledge regarding this sub-concept, as can be seen from the higher test scores in the pre test for both groups. As already explained in the theoretical background, ICL is co-determined by the prior knowledge of the students, so that it can be assumed that ICL for learning tasks concerning this sub-concept was lower for the subject area uniform motion than for the subject area accelerated motion. It can also be assumed that ECL only has a significant influence on learning performance when ICL is high (Paas et al 2003; Zheng 2018). It can be concluded that the load-reducing effect of video analysis only in the subject area of accelerated motion could have had a significant influence on learning performance with respect to this sub-concept since ICL was higher than in the subject area of uniform motion.

In summary, the positive effects of the preliminary study regarding the conceptual understanding could be replicated for two sub-concepts, and the learning-promoting effect could be attributed to the interactivity of the video analysis application as well as the dynamic linking of coordinate system and motion diagrams. This supports research findings that have already empirically proven a positive effect of video analysis on conceptual understanding for other topics in mechanics (Hochberg et al 2020; Wee et al 2015; Hockicko et al 2014).

Intervention-induced Cognitive Load

By using the test instrument developed by Leppink et al (2013), it could be demonstrated that the extraneous cognitive load for the technology-supported experiment-based learning process is significantly lower for both subject areas. Thus, the theoretically derived argumentation that the visualization of multiple forms of representation by means of the video analysis application fulfills the design principles for multimedia learning environments and that this leads to a reduction of the extraneous load is empirically supported. In addition, it can be noted that for the subject area of accelerated motion, the learners of the TG estimated the intrinsic load to be significantly lower. Since the prior knowledge of the learners also contributes to the intrinsic load, and the experiments were identical, and the learning tasks were comparable for both groups, a possible explanation could lie in the higher prior knowledge of the learners of the TG regarding sub-concept G3 of uniform motion. Furthermore, the previous experiences from the execution

of the experiments on uniform motion might have influenced the learners' assessment, so that the students of the TG found the learning tasks on accelerated motion less complex than did the students of the CG.

Intervention-induced Emotions

It could be shown that the emotional state of the learners differs significantly between technology-supported and traditional experiment-based learning processes. Thus, technology support leads to a weaker formation of negative-deactivating emotions. This consistently extends the research results of Hillmayr et al (2020) and Sung et al (2016), which could prove a positive influence on motivational variables in technology-supported learning environments by the positive effect on emotional variables. However, it should be noted that the positive effect is limited to negative-deactivating emotions. In contrast, no significant group difference could be demonstrated for positive-activating emotions.

Cause-effect Relationships Between Emotions, Cognitive Load, and Learning Achievement

For both subject areas, a positive cause-effect relationship between negative-deactivating emotions and extraneous cognitive load could be statistically proven by structural equation modeling. Thus, the greater the formation of negative emotions during the experiment-based learning process, the greater the extraneous load. This is in accordance with the ICALM model, which postulates an influence of negative emotions on the learning process by additionally binding cognitive resources, and thus, impairing the active knowledge construction. This assumption is also supported by current research results (Brand et al 2007; D'Mello and Graesser 2012).

For positive-activating emotions, no influence on cognitive load could be proven. According to Pekrun (2014), positive emotions support the use of flexible learning strategies and self-regulated learning, which implies a positive influence on cognitive load. One possible explanation for the absence of this influence is the strong control of the intervention. To ensure a fair comparison between TG and CG, the learning process was pre-structured for the students and left no room for other learning strategies so that there was no need for self-regulatory mechanisms for the students.

For both topics, a negative cause-effect relationship between extraneous cognitive load and learning achievement was found for the sub-concepts for which a positive effect in favor of the TG was identified. In conclusion, the reduction of extraneous cognitive load as theoretically hypothesized can be regarded as a cause for the increased learning achievement. A direct influence of teacher behavior on learning achievement could not be found, which in-

dicates that the positive effects found are independent of the instructing teacher. However, an influence on the extraneous load was detected for the uniform motion. A possible explanation could be that the students in this subject area are using tablet-supported video analysis for the first time independently when experimenting. A lack of support from the teacher seems to increase the extraneous load of the students, although the teachers took a passive role in the learning process. This is also supported by the fact that in the following subject area, where the students were already familiar with the measurement methodology, no influence of the teacher's behavior could be proven.

Conclusion

With regard to the ongoing digitization, especially in the field of science education, this work aimed at contributing to closing the large research gap regarding replicable results on the learning effectiveness of digital learning tools in real-life teaching scenarios. Thus, it was possible to replicate the positive effects of the use of tablet-supported video analysis in a replication study for two essential topics in the teaching of mechanics, uniform, and accelerated motion. Increasing the sample size compared to the preliminary study could increase the statistical power, contributing to the generalization of the results. The effects that could be replicated in this way were demonstrated in both the preliminary and replication study with high test power, which indicates that the occurrence of these effects is independent of the sample population. Using multilevel regression analysis could also show that the course affiliation, and thus, the instructing teacher, has only a minor influence on the effects observed. This also indicates the generalizability of the replicated effects. This work thus contributes to the transfer of the findings into school practice and to the evidence-based implementation in regular physics lessons. It should be noted that for both extraneous load and negative-deactivating emotions, the proportion of explained variance at course level is higher for uniform motion than for accelerated motion. One possible explanation for this difference is that at the beginning of the intervention on the subject of uniform motion, students are confronted with a new type of measurement methodology in the classroom and are therefore uncertain at first what to expect. In this situation, the teacher's behavior may have a greater influence on the emotional state of the learners and thus the cognitive load than in the teaching sequence for accelerated motion, in which the students are already familiar with the measurement methodology.

A second objective of the replication study was to empirically support the theory-based foundation of the learning effectiveness in terms of the reduction of cognitive load. It was possible to demonstrate with high test power that the extraneous cognitive load during the experiment-based learning process in the technology-supported teaching sequences is significantly lower than in the traditional teaching sequences. By means of structural equation modeling, the theoretically hypothesized direct causal relationship between load reduction and learning achievement could be empirically verified. This supports the theory-guided argumentation that the learning-promoting effect of the digital tool results from the reduction of the load irrelevant to learning. The learner thus has more free cognitive resources available for active knowledge construction, which increases the effectiveness of the learning process and ultimately leads to a deeper conceptual understanding. This supports the theoretical foundation of effectiveness of tablet-based video analysis derived from multimedia learning theories and contributes to an expansion of the research basis by providing a possible explanation for the positive effects on learning achievement already proven in several studies (Becker et al 2018, 2019; Hochberg et al 2020; Klein et al 2018). Furthermore, it could also be shown that basic theoretical design principles for multimedia learning environments can also have a learning-promoting effect in real multirepresentational teaching scenarios.

Additionally, it was shown that the use of technology also has an impact on the emotional state of the learners. Thus, learners who were supported by technology when experimenting showed a significantly lower level of negative-deactivating emotions. Using structural equation modeling statistically proved that, in accordance with the ICALM model, a higher level of negative emotions leads to an increase in extraneous load. The study results thus reveal an indirect learning effect on the influence of digital learning tools on the emotional state, which has so far received little attention, and contribute to closing the large gap between the ubiquity of emotions in learning processes and the focus of educational research on mostly cognitive variables, which exists according to Pekrun and Stephens (2010) and Plass and Kaplan (2016).

Outlook

Due to the product orientation of the study, the question remains open to what extent and at what point in the learning process the learners use the available forms of representation individually or in combination. To clarify this question, a process-oriented research methodology such as eye tracking is suitable. This would make it possible to capture the interaction of the learners with the forms of representation provided by the video analysis application with a high temporal resolution. In addition to the allocation of attention to certain forms of representation, transitions between them

could also be investigated, which would provide insight
into the integration of different forms of representation in
the solving of a physical problem using the digital tool. In
this way, the learner's use of the dynamic link between the
coordinate system and the motion diagrams could also be
resolved, which is a possible reason for the learning effec-
tiveness of the digital tool. Theoretical models of the effec-
tiveness of technology support in learning with MERs (e.g.,
Rau 2017) could thus be empirically supported or even ex-
tended. Against the background of a not yet completed ex-
tension of traditional multimedia theories to experiment-
based learning processes, the process-oriented investigation
of the interaction of learners with MERs is a promising re-
search desideratum.

Furthermore, based on the study findings, no statement
can be made about the effectiveness of long-term use in
regular school lessons. However, due to the comparable ef-
fectiveness for two chronologically consecutive, differently
complex subject areas, it can be assumed that the positive
effects of this method for more complex topics in the teach-
ing of mechanics will at least be maintained, which shows
the great potential of tablet-supported video analysis for
sustainable use in physics lessons. In view of the ongoing
digital transformation in education, we encourage other re-
search groups to build on the results of the study presented
here and to explore the potential of this digital learning
tool for other (more complex) topics in mechanics as well
as a sustainable long-term use in regular physics lessons.

References

Ainsworth, S. (2006). Deft: a conceptual framework for considering
    learning with multiple representations. *Learning and Instruction*,
    16(3), 183–198.
Ainsworth, S. (2008). *The educational value of multiple-representa-
    tions when learning complex scientific concepts* (pp. 191–208).
    Dordrecht: Springer.
Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error
    on convergence, improper solutions, and goodness-of-fit indices
    for maximum likelihood confirmatory factor analysis. *Psychome-
    trika*, *49*(2), 155–173.

Barrett, P. (2007). Structural equation modelling: adjudging model fit.
    Personality and Individual Differences, *42*(5), 815–824.
Becker, S., Klein, P., Gößling, A., & Kuhn, J. (2019). Förderung von
    Konzeptverständnis und Repräsentationskompetenz durch Tablet-
    PC-gestützte Videoanalyse. *Zeitschrift für Didaktik der Naturwis-
    senschaften*, *54*(7), 1–24. Promoting conceptual understanding
    and representational competence through tablet PC-based video
    analysis.
Becker, S., Klein, P., Gößling, A., & Kuhn, J. (2020). Using mobile de-
    vices to enhance inquiry-based learning processes. *Learning and
    Instruction*, *69*, 101350.
Becker, S., Klein, P., & Kuhn, J. (2018). Promoting students' concep-
    tual knowledge using video analysis on tablet computers. In A.
    Traxler, Y. Cao & S. Wolf (Eds.), *PERC Proceedings*.
Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on
    the fragility of performance: choking under pressure in mathemat-
    ical problem solving. *Journal of Experimental Psychology: Gen-
    eral*, *133*(4), 584–600.
Brand, S., Reimer, T., & Opwis, K. (2007). How do we learn in a neg-
    ative mood? Effects of a negative mood on transfer and learning.
    Learning and Instruction, *17*(1), 1–16.
diSessa, A. A. (2004). Metarepresentation: Native competence and tar-
    gets for instruction. *Cognition and Instruction*, *22*(3), 293–331.
D'Mello, S., & Graesser, A. (2012). Dynamics of affective states dur-
    ing complex learning. *Learning and Instruction*, *22*(2), 145–157.
Fan, X., & Nowell, L. D. (2011). Using propensity score matching in
    educational research. *Gifted Child Quarterly*, *55*(1), 74–79.
Ginns, P. (2006). Integrating information: a meta-analysis of the spatial
    contiguity and temporal contiguity effects. *Learning and Instruc-
    tion*, *16*(6), 511–525.
Guilford, J. P. (1982). Cognitive psychology's ambiguities: some sug-
    gested remedies. *Psychological Review*, *89*(1), 48–59.
Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: statistical
    methods and applications*. Thousand Oaks: SAGE.
Haagen-Schützenhöfer, C., & Joham, B. (2018). Professionalising
    physics teachers in doing experimental work. *CEPS Journal*,
    8(1), 9–34.
Hadie, S. N. H., & Yusoff, M. S. B. (2016). Assessing the validity of the
    cognitive load scale in a problem-based learning setting. *Journal
    of Taibah University Medical Sciences*, *11*(3), 194–202.
Halloun, I. A. (2006). *Modeling theory in science education*. Rotter-
    dam: Springer.
Harp, S. F., & Mayer, R. E. (1998). How seductive details do their dam-
    age: a theory of cognitive interest in science learning. *Journal of
    Educational Psychology*, *90*(3), 414–434.
Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000).
    What is the purpose of this experiment? Or can students learn
    something from doing experiments? *Journal of Research in Sci-
    ence Teaching*, *37*(7), 655–675.
Haury, D. L., & Rillero, P. (1994). *Perspectives of hands-on science
    teaching*. Columbus, OH: ERIC Clearinghouse for Science, Math-
    ematics and Environmental Education.
Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M.
    (2020). The potential of digital tools to enhance mathematics and
    science learning in secondary schools: A context-specific meta-
    analysis. *Computers & Education*, *153*, 103897.
Hochberg, K., Becker, S., Louis, M., Klein, P., & Kuhn, J. (2020).
    Using smartphones as experimental tools—a follow-up: cognitive
    effects by video analysis and reduction of cognitive load by mul-
    tiple representations. *Journal of Science Education and Technol-
    ogy*, *29*(2), 303–317.
Hochberg, K., Kuhn, J., & Müller, A. (2018). Using smartphones as
    experimental tools—effects on interest, curiosity, and learning in
    physics education. *Journal of Science Education and Technology*,
    27(5), 385–403.
Hockicko, P., Trpišová, B., & Ondruš, J. (2014). Correcting students'
    misconceptions about automobile braking distances and video

analysis using interactive program tracker. *Journal of Science Education and Technology, 23*(6), 763–776.

Horz, H., Schnotz, W., Plass, J.L., Moreno, R., & Brünken, R. (2009). *Cognitive load in learning with multiple representations* (pp. 229–252). Cambridge: Cambridge University Press.

Hox, J.J. (2010). *Multilevel analysis: techniques and applications* (2nd edn.). New York: Routledge/Taylor & Francis Group.

Hox, J.J., & McNeish, D. (2020). *Small samples in multilevel modeling* (pp. 215–225). London, New York: Routledge.

Husnaini, S.J., & Chen, S. (2019). Effects of guided inquiry virtual and physical laboratories on conceptual understanding, inquiry performance, scientific inquiry self-efficacy, and enjoyment. *Physical Review Physics Education Research, 15*(1), 10119.

Johnstone, A., & Wham, A. (1982). The demands of practical work. *Education in Chemistry, 19*(3), 71–73.

de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J., van Someren, M., Spada, H., & Swaak, J. (1998). *Acquiring knowledge in science and mathematics: the use of multiple representations in technology based learning environments*. Advances in learning and instruction series. (pp. 9–40). Amsterdam: Pergamon.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*(4), 443–477.

Kapici, H.O., Akcay, H., & de Jong, T. (2019). Using hands-on and virtual laboratories alone or together–which works better for acquiring knowledge and skills? *Journal of Science Education and Technology, 28*(3), 231–250.

Kenward, M.G., & Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997.

Klein, P., Kuhn, J., & Müller, A. (2018). Förderung von Repräsentationskompetenz und Experimentbezug in den vorlesungsbegleitenden Übungen zur Experimentalphysik. *Zeitschrift für Didaktik der Naturwissenschaften, 24*(1), 17–34. Promotion of representational competence and experiment reference in the lecture accompanying exercises on experimental physics.

Kline, R.B. (2011). *Principles and practice of structural equation modeling* (3rd edn.). New York: Guilford.

Kuhn, J., & Vogt, P. (2015). *Smartphones & Co. in physics education: effects of learning with new media experimental tools in acoustics* (pp. 253–269). Basingstoke, UK: Palgrave Macmillan.

Leppink, J. (2017). Cognitive load theory: practical implications and an important challenge. *Journal of Taibah University Medical Sciences, 12*(5), 385–391.

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–127.

Leppink, J., Paas, F., van der Vleuten, C.P.M., van Gog, T., & van Merriënboer, J.J.G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072.

Lindsay, R.M., & Ehrenberg, A.S.C. (1993). The design of replicated studies. *The American Statistician, 47*(3), 217–228.

Luke, S.G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*, 1494–1502.

Makel, M.C., & Plucker, J.A. (2014). Facts are more important than novelty. *Educational Researcher, 43*(6), 304–316.

Mayer, R.E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research, 31*, 611–624.

Mayer, R.E. (2005). *Cognitive theory of multimedia learning* (pp. 31–48). Cambridge: Cambridge University Press.

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

McDermott, L.C., Rosenquist, M.L., & van Zee, E.H. (1987). Student difficulties in connecting graphs and physics: examples from kinematics. *American Journal of Physics, 55*(6), 503–513.

McNeish, D.M., & Stapleton, L.M. (2016). The effect of small sample size on two-level model estimates: a review and illustration. *Educational Psychology Review, 28*(2), 295–314.

van Merriënboer, J.J.G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177.

van Meter, P., List, A., Lombardi, D., & Kendeou, P. (2020). *Handbook of learning from multiple representations and perspectives*. New York: Routledge.

Moreno, R. (2005). *Instructional technology: promise and pitfalls* (pp. 1–19). Greenwich, CT: Information Age Publishing.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158.

Moreno, R., & Mayer, R.E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

Moreno, R., & Mayer, R.E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326.

Muthén, L.K., & Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural equation modeling, 9*(4), 599–620.

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: a systematic review. *Computers & Education, 141*, 103618.

Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force concept inventory-based multiple-choice test for investigating students' representational consistency. *Physical Review Physics Education Research, 6*(2), 020,109–020,112.

Oaksford, M., Morris, F., Grainger, B., & Williams, J.M.G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 476–492.

Oliveira, A., Behnagh, R.F., Ni, L., Mohsinah, A.A., Burgess, K.J., & Guo, L. (2019). Emerging technologies as pedagogical tools for teaching and learning science: a literature review. *Human Behavior and Emerging Technologies, 1*(2), 149–160.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*(1), 1–4.

Pekrun, R. (2000). *A social-cognitive, control-value theory of achievement emotions* (pp. 143–163). New York: Elsevier.

Pekrun, R. (2014). *Emotions and Learning. Educational Practices Series*, vol. 24. Geneva, Switzerland: International Academy of Education (IAE) and International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Pekrun, R., & Stephens, E.J. (2010). Achievement emotions: a control-value approach. *Social and Personality Psychology Compass, 4*(4), 238–255.

Pekrun, R., Frenzel, A.C., Barchfeld, P., & Perry, R.P. (2011). Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36–48.

Pekrun, R., Götz, T., Titz, W., & Perry, R.P. (2002). Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–106.

Plass, J.L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*(2), 339–359.

Plass, J.L., & Kaplan, U. (2016). *Emotional design in digital media for learning* (pp. 131–161). San Diego: Academic Press.

Rau, M.A. (2017). A framework for educational technologies that support representational competencies. *IEEE Transactions on Learning Technologies, 10*(3), 290–305.

Robinson, W.R. (2004). Cognitive theory and the design of multimedia instruction. *Journal of Chemical Education, 81*(1), 10.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rost, D. H., & Bienefeld, M. (2019). Nicht replizieren: publizieren!? *Zeitschrift für Pädagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000253. Not replicate: publish!?.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*(2), 227–237.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton, Mifflin and Company.

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: a meta-analysis and research synthesis. *Computers & Education*, *94*, 252–275.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*, 257–285.

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

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*. https://doi.org/10.1007/s10648-019-09465-5.

Treagust, D., Duit, R., & Fischer, H. (2017). *Multiple representations in physics education*. Dordrecht: Springer.

Tytler, R., Prain, V., Hubber, P., & Waldrip, B. (2013). *Constructing representations to learn in science*. Rotterdam: SensePublishers.

Verschaffel, L., de Corte, E., de Jong, T., & Elen, J. (2010). *Use of representations in reasoning and problem solving*. New York: Routledge.

Volkwyn, T. S., Allie, S., Buffler, A., & Lubben, F. (2008). Impact of a conventional introductory laboratory course on the understanding of measurement. *Physical Review Special Topics – Physics Education Research*, *4*(1), 108–10.

Wee, L. K., Tan, K. K., Leong, T. K., & Tan, C. (2015). Using Tracker to understand toss up and free fall motion: a case study. *Physics Education*, *50*(4), 436–442.

Woolnough, B. E. (1979). The role of the laboratory in physics education. *Physics Education*, *14*(2), 70–74.

Zheng, R. Z. (Ed.). (2018). *Cognitive load measurement and application: a theoretical framework for meaningful research and practice*. New York: Routledge/Taylor & Francis.

Zukić, M., Đapo, N., & Husremović, D. (2016). Construct and predictive validity of an instrument for measuring intrinsic, extraneous and germane cognitive load. *Universal Journal of Psychology*, *4*(5), 242–248.

Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: review of research. *Computers & Education*, *94*, 1–17.