preprint

Superfrog: Comparing Learning Outcomes and Potentials of a Worksheet, Smartphone, and Tangible AR Learning Environment

Sebastian Oberdörfer[1][0000-0002-8614-1888], Anne Elsässer[1], Silke Grafe[2][0000-0003-3970-0859], and Marc Erich Latoschik[1][0000-0002-9340-9600]

[1] Human-Computer Interaction, University of Würzburg, Würzburg, Germany
[2] School Pedagogy, University of Würzburg, Würzburg, Germany

Abstract. The widespread availability of smartphones facilitates the integration of digital, augmented reality (AR), and tangible augmented reality (TAR) learning environments into the classroom. A haptic aspect can enhance the user's overall experience during a learning process. To investigate further benefits of using TAR for educational purposes, we compare a TAR and a smartphone learning environment with a traditional worksheet counterpart in terms of learning effectiveness, emotions, motivation, and cognitive load. 64 sixth-grade students from a German high school used one of the three conditions to learn about frog anatomy. We found no significant differences in learning effectiveness and cognitive load. The TAR condition elicited significantly higher positive emotions than the worksheet, but not the smartphone condition. Both digital learning environments elicited significantly higher motivation, in contrast to the worksheet. Thus, our results suggest that smartphone and TAR learning environments are equally beneficial for enhancing learning.

Keywords: Augmented Reality · Education · Serious Games · Gamification · Tangible User Interfaces.

1 Introduction

The widespread availability of smartphones allows teachers to develop technology-based teaching concepts [20]. Using smartphones not only provides access to digital learning environments, but also enables an integration of Augmented Reality (AR) in teaching concepts [56]. Following an approach of media didactics, a learning environment provides a simulation of a given subject that enables learners to interact as well as experiment with and to observe the results of their actions [53]. Thus, a learning environment requires self-directed learning and, depending on the desired structure of the learning, finding a solution to a problem, design of a product, and evaluation of a situation. Using AR for teaching and learning can result in higher learning gains, motivation, and experiential gains through direct application of learning content [4].

AR-based learning environments can also be extended by a tangible user interface (TUI), thus giving the learning process a physical aspect [8]. Combining

preprint

Fig. 1. When detecting the markers with a smartphone, *Horst-TAR* displays 3D models of the organs. Touching the organ reveals biological information.

physical objects with so-called fiducial markers achieves *Tangible AR* (*TAR*) [8]. TAR allows learners to inspect augmented objects from all angles by manipulating them directly in six degrees of freedom [5]. This intensifies the direct application of learning content and enables spatial learning.

TAR results in a higher complexity of the learning environment. Besides the requirement of a smartphone, a teacher also must provide enough physical markers for the entire class. This becomes even more challenging when learners should continue to learn at home. In addition, teachers must acquire technological pedagogical content knowledge to successfully integrate AR learning environments in classroom teaching [28, 39]. Hence, the educational benefits of using TAR should outweigh the preparation complexity to justify a classroom integration. Therefore, it is of high importance to investigate the potential advantages of TAR for educational purposes to derive guidelines as well as recommendations for developers and educators.

Contribution

The present study investigates the learning effectiveness of a TAR learning environment in comparison to a smartphone and traditional worksheet counterpart at a local high school. We use the TAR and smartphone version of the gamified learning environment *Horst – The Teaching Frog* [41]. To enable a baseline measurement, we designed a worksheet providing the same declarative information about a frog's organs. The learning environments target the learning of the anatomy of frogs as shown in Fig. 1.

While the digital versions simulate dissection, the worksheet approach conveys the information with illustrations and text only. The three versions of the learning environment differ only in the way knowledge is presented, i.e., smartphone, TAR, and worksheet. In our user study, we found no significant difference between the conditions in terms of learning effectiveness and cognitive load. However, the TAR version elicited significantly higher positive emotion than the worksheet, but not the smartphone version. In addition, the TAR and smartphone versions elicited significantly higher motivation compared to the

preprint

worksheet. Thus, our results suggest that smartphone and TAR learning environments are equally beneficial for enhancing learning, with TAR even eliciting slightly higher positive emotion.

2 Theoretical Background

AR three-dimensionally integrates virtual elements into the real-world that are interactive in real time [2]. Users can experience AR using headworn, handheld, and projected displays [27]. Handheld AR commonly is achieved using smartphones and tablets that are turned into a "Magic Lens" [6] revealing the augmentations. However, headworn and projected displays provide a greater freedom to users by keeping their hands free. Smartphone AR benefits from the widespread availability and familiarization of the users with these devices [32]. This makes smartphone AR especially useful for educational use cases [56].

Besides interacting with virtual elements on the device, they can also be manipulated using a TUI [19]. A TUI uses real world objects as input and output devices, thus connecting digital information with real world objects [23]. A TAR interface similarly links virtual information to physical objects. This allows for an augmentation of physical objects and an interaction with the AR system by manipulating the respective objects [8]. TAR not only suits the visualization of 3D models [5], but also yields a very intuitive experience [8].

2.1 Benefits of AR for Learning

Using AR for educational purposes can facilitate the learning and knowledge acquisition [10]. This potentially can lead to an overall positive attitude towards the learning content and hence to higher academic success [24]. AR learning environments further assist the learning of complex constructs by providing spatial and direct visualization of learning content [15]. For instance, *Mathland* demonstrates the mathematics behind Newtonian physics and allows users to modify and hence explore the physical laws [26]. However, AR can also be beneficial for vocabulary learning leading to better short-term retention in contrast to a non-AR counterpart [54]. Finally, as AR allows for a direct interaction with real-world objects, AR learning environments can further support the requirements of special needs education [49]. Similarly, a TAR learning environment can evoke a higher degree of joy and motivation in comparison to Graphical User Interfaces (GUIs) [22] as well as reduce the cognitive load [12]. Using TAR in an educational context can intensify work on learning material, improve usability, and support mental skills as well as collaboration [1]. Overall, TAR learning environments address three core aspect particularly relevant for learning according to the theories of self-determination [47] and cognitive load [51]: positive emotions, higher motivation, and lower cognitive load. These are also key attributes of learning using extended reality [13].

Different emotions can have different effects on the mediators memory processes, regulation of learning, cognitive load, and motivation [45]. Thus, it is

preprint

assumed that learning performance is positively influenced by positive emotions like joy and negatively influenced by emotions with negative valence, such as anxiety [33]. Zuckerman and Gal-Oz [58] showed a higher preference and a higher rating with respect to stimulation and entertainment of TUIs in comparison to GUIs. Oberdörfer et al. [40] found first indications that TAR learning environments are more stimulating, attractive, and novel as well as yield a higher preference in comparison to AR and smartphone counterparts.

Intrinsic motivation or internalized forms of extrinsic motivation lead to higher learning performance than externally regulated forms of motivation [46]. AR demonstrated to evoke such a higher degree of motivation in comparison to other learning media [18]. For example, students experienced high motivation in learning about georeferenced information when using an AR tool to visualize relief [11]. This effect could be due to the perceived autonomy in using the technology and the multisensory experience that allows learners to interact in a natural way [18].

The goal of successful instructional design is to make optimal use of the capacity of working memory. To achieve this, multimedia principles can be implemented to reduce extraneous cognitive load, manage intrinsic cognitive load, and promote germane cognitive load [35]. Digital learning environments can present and allow for a direct interaction with three-dimensional information. In contrast to printed textbooks, this reduces the cognitive load of mentally rotating objects to analyze and understand them. For instance, learners reported lower extraneous and intrinsic cognitive load when completing chemistry exercises [25] or learning anatomy [29] in AR.

2.2 Playful Learning Experiences

The overall learning process can be embedded in a gamified approach [38]. Gamified learning environments can either be *serious games* [17] or non-gaming learning applications enhanced by *gamification* [48]. Following such an approach, the learning can become an engaging, vivid, and inspiring experience [36]. Serious games map the learning contents to central game mechanics or core interactions, thus achieving their application and demonstration [42]. Gamification refers to the integration of game elements in non-gaming environments [14]. Gamification enhances the motivation of pupils [34] leading to more repetitive learning and development of sustainable knowledge. This general effect of gamification also applies to AR learning environments, yielding a higher joy, interest, and engagement [31].

3 Horst – The Teaching Frog

To investigate the effects of TAR technology on learning, we selected the gamified TAR learning environment *Horst – The Teaching Frog* [41]. *Horst – The Teaching Frog* simulates a dissection of a frog, thus enabling the learning about a frog's anatomy. Digital simulations can yield a more effective learning in comparison to

preprint

real dissections [57]. The learning environment is designed as a supplementary material for sixth grade biology lessons dealing with anatomy of amphibians [7]. During the simulation, learners can either acquire and deepen the encoded knowledge in a virtual dissection or assess their learning progress in a quiz. The learning content presented is based on two textbooks on amphibians [21, 50] and defined in collaboration with educators. Besides the TAR version (*Horst-TAR*), a smartphone (*Horst-S*) and AR-only version (*Horst-AR*) were developed that encode the same learning content, provide the same functions, but differ in the technology used [40]. A study revealed that all versions of Horst are comparable with respect to intuitive use, but differ regarding user experience [40]. User experience was highest for *Horst-TAR* followed by *Horst-AR*. Thus, this study paved the way for our experiment by ruling out potential confounds on our learning effectiveness measurements caused by differences in the usability.

To compare the TAR learning environment to the de-facto standard of digital learning environments, i.e., smartphones, we also included *Horst-S* in our study. We further designed an additional worksheet version (*Horst-WS*) to compare *Horst-TAR* to a traditional learning method and to generate a baseline measurement. *Horst-WS* presents the same learning content as the digital versions and follows the structure of common textbooks. This ensures that measured differences in the learning outcome are caused by the technology and not differences in the learning contents. However, we decided against including *Horst-AR* in our experiment. Similar to TAR, AR would cause a higher preparation complexity as additional markers must be provided besides AR devices.

Horst – The Teaching Frog is available for download at our lab's website[3].

3.1 Digital Learning Environments

The digital versions of *Horst* provide two dissection modes, i.e., assisted and free dissection, a quiz to self-assess the individual learning progress, and an achievement as well as highscore system. The assisted mode guides a user through the process of dissecting a frog, thus scaffolding the learning process. By explaining each step in detail, this mode provides additional information about the anatomy of a frog. After reading the description of a step provided by the pedagogical agent Horst, learners must find the relevant organ, extract it, and display as well as read the organ's biological information as displayed in Fig. 1. Subsequently, the learning environment displays the next task. Learners cannot skip a step during the assisted dissection. In contrast, the free dissection provides no guidance, but allows for a free examination of the frog's organs. The learning environment currently includes seven organs that are sequenced in the following order during the assisted mode: heart, liver, lungs, stomach, gut, kidney, and bladder.

Horst-TAR is based on a large, but realistic soft toy of a frog. A pouch featuring a zipper was added to the belly of the plush frog, thus allowing for its dissection as shown in Fig. 1. Inside this pouch, extractable *tangible* paper-card-based markers are attached to the frog using a piece of velcro. Each tangible

[3] https://hci.uni-wuerzburg.de/projects/horst-the-teaching-frog/

preprint

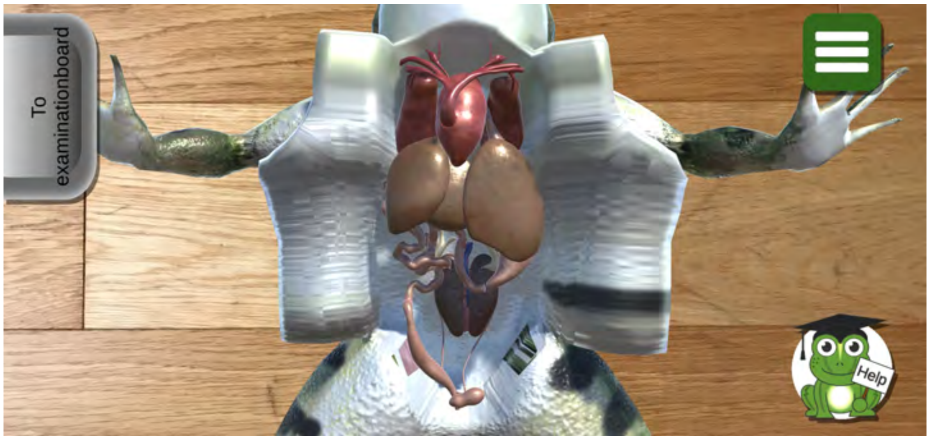

Fig. 2. *Horst-S* allows for a frog dissection on smartphones, thus presenting the defacto standard for mobile learning.

object represents an individual organ and features two 2D images of the respective organ as image targets. The front side shows an image of the organ's outer side and the back side an image the organ's inner side as seen from the belly. The *Horst-TAR* application displays a 3D model of each organ above the relevant markers when detecting them with the smartphone's camera. This visualizes the anatomy of a frog. By physically extracting a marker, learners can inspect the 3D organ model from all angles up close. While scanning the front side, learners can inspect the outer side of the organ. Vice-versa, scanning the back side allows for an inspection of the organ's inner side. *Horst-TAR* displays biological information about an organ when touching it on the smartphone display. In contrast, *Horst-S* displays a 3D model of a frog and its individual organs as shown in Fig. 2. A user merely needs to touch an organ to display the relevant biological information and to inspect it up close using drag gestures.

The quiz contains 16 multiple-choice questions that test learners' knowledge of frog anatomy and functions of their organs. Learners receive an immediate audiovisual feedback about the correctness of their selections. While correct answers are marked in green and rewarded with a quack sound, wrong answers are marked in red and emphasized with the sound of a buzzer. The learning environment removes a correctly answered question from the list but returns a wrongly answered question to it. In this way, learners get an additional chance to reflect about the exercise and hence to deepen their knowledge.

Finally, the learning environment motivates the learning process and repetition of the provided learning opportunities with an achievement and highscore system. Achievements present clear tasks like completing five dissections, thus giving learners an incentive to repeat the process. The highscore system rewards the completion of an assisted dissection as well as a good performance in the quiz.

preprint


Frog organs and their functions

Fig. 3. We designed a worksheet version for the learning content presented in the digital versions.

3.2 Worksheet

Horst-WS consists of a short instruction, a picture of the 3D-model used in *Horst-S* supplemented by the labeling of the individual organs, and information texts on the individual organs in dissection order as displayed in Fig. 3. For the purpose of contextualization, we added an adapted version of the introductory text of the chapter "The water frog – a life in water and on land" from the biology textbook Natura 1 [3] to the worksheet and illustrated it with two photos of a frog. The introductory text is also used in the digital versions, but narrated by the agent Horst from the first-person perspective. Thus, the three versions of *Horst – The Teaching Frog* only differ with respect to the presentation technology used.

4 Study Design

The overall goal of our research is to investigate the effects of using TAR in comparison to smartphones and traditional worksheets for educational purposes. In particular, we targeted the acquisition of declarative knowledge about the organs and their functions. We defined the following learning goal: *After the learning phase, the participants can name and explain the seven included organs.*

To ensure for such an acquisition of declarative knowledge and hence a comparability of our results, we limited the functions of *Horst-TAR* and *Horst-S* to the assisted dissection, only. Also, we disabled the quiz as well as highscore feature and reduced the achievement system to a minimum. Participants receive four achievements throughout the entire learning process. The first one is unlocked when starting the tutorial, the second one when finishing the tutorial, the third one after having inspected the first half of the organs, and the last one

preprint

upon completion of the dissection. In this way, the three conditions focused the memorization and understanding of facts.

We assume the following hypotheses based on the analysis of the theoretical work in Section 2 and the design of *Horst – The Teaching Frog* in Section 3.

H1 The learning effectiveness is higher with *Horst-TAR* than *Horst-S* and *Horst-S* than *Horst-WS*.
H2 A learner's emotions are more positive after using *Horst-TAR* than *Horst-S* and *Horst-S* than *Horst-WS*.
H3 A learner's motivation is higher when learning with *Horst-TAR* than *Horst-S* and *Horst-S* than *Horst-WS*.
H4 A learner's cognitive load is lower when learning with *Horst-TAR* than *Horst-S* and *Horst-S* than *Horst-WS*.

To compare the three versions and to answer the four hypotheses, we conducted a user study following a between-groups design. Participants were randomly assigned to either one of the three learning environments and completed an assisted dissection. Before and after the learning phase, participants answered various questionnaires. In addition, we conducted a semi-structured interview with teachers to gain additional insights into the feasibility of integrating TAR learning environments in classroom teaching.

The institutional review board of Human-Computer Media at the University of Würzburg approved our ethics proposal for this study.

4.1 Measures

We used the following measures to compare the conditions.

Learning Effectiveness To measure learning effectiveness and to factor out a potential influence of prior knowledge, we assessed the participants' knowledge with a written exam before and after learning. Both exams tested our defined learning goal, i.e., knowing and comprehending the seven included organs, and contained the same 10 exercises. We presented the exercises in randomized order except for the first task, i.e., labeling the organs. The remaining exercises required the participants to either name the organ that fulfills a specific purpose or to describe functions of a particular organ. The formulation of the exercises was based on official guidelines and reviewed by biology teachers. The maximum achievable score is 22.

The exam consisted of the following exercises:

1. Labeling of the individual organs depicted in a screenshot of the 3D model used in *Horst-S*.
2. Name the scientific term for the intestinal tract of frogs.
3. Name the scientific term for the mixture of oxygenated and deoxygenated blood.
4. Name the two organs that are part of the digestive tract.

preprint

5. Name the organ that pumps blood through the frog's body.
6. Name the breathing type of frogs in which they take in oxygen through their skin.
7. Name the two functions of the liver.
8. Name the parts into which the heart is divided. Use scientific terms.
9. Describe how oral cavity breathing works in frogs.
10. Name the two functions of the kidney.

Emotion We assessed the current affect using a short version of the *Positive and Negative Affect Schedule for Children (PANAS-C)* [16]. The short version of the PANAS-C contains 10 items, which can be equally divided into the dimensions Positive Affect (PA) and Negative Affect (NA). We sequenced the items according to the original version of the PANAS-C [30]. We directly translated each item to German, e.g., "sad" to "traurig" and "joyful" to "fröhlich", and used the validated German translation of the Likert scale labels [9]. The instructions were also based on the validated German version of the *PANAS* [9], but adjusted in their wording to facilitate understanding for the children.

Motivation To measure motivation, we used the *Short Scale of Intrinsic Motivation (KIM)* [55]. The scale is based on the *Intrinsic Motivation Inventory (IMI)* [47] and was developed for high school students. It contains 3 items each for the factors interest/enjoyment (short: enjoyment), perceived competence (short: competence), perceived choice (short: choice), and pressure/tension (short: pressure), and thus a total of 12 items. In the original, the 5-point Likert scale ranges from 0 "do not agree at all" to 4 "agree completely". We changed the numbering to 1 "do not agree at all" to 5 "agree completely" for the purpose of uniform interpretation.

Cognitive Load We used the 9-point *Paas* subjective rating scale [43, 44] ranging from 1 "very, very little mental effort" to 9 "very, very high mental effort" to assess the experienced cognitive load. Although the scale does not differentiate between extraneous and intrinsic of cognitive load, the scale has been shown to be a valid measurement tool for cognitive load in general and requires little time to complete [52]. We translated the scale to German and reworded it for improving comprehensibility by directly naming the learning content: "In learning the organs of the frog, I invested ..." / "Beim Lernen der Organe des Frosches habe ich mich ... angestrengt".

Feedback We posed three closed questions to obtain subjective feedback from the participants on their version of the learning environment and respective learning process. Participants could agree to statements about the willingness to repeat learning with their learning environment and desire to learn with the respective technology in other subjects. We used the same 5-point Likert scale as in the KIM. In addition, we asked the participants to rate their learning environment with school grades ranging from 1 "very good" to 6 "unsatisfactory".

preprint

Teacher Interview To gain preliminary feedback on the potential integration of *Horst – The Teaching Frog* in teaching concepts, we gauged the teachers' perspectives on the different learning environments. We conducted an oral semi-structured interview with three teachers of whom two were pre-service teachers. At the end of a study day, the teachers and experimenter came together to inspect the learning environments. Before allowing the teachers to use the learning environments, we gave them a quick overview and explained all possible interactions. Subsequently, we conducted the interview. In favor of the flow of speech, we did not pay attention to a consistent wording of the questions. The teacher interview took place in parallel to our user study. The question pool included the following items:

1. Which of the three learning environments do you prefer and why?
2. How would you use your favorite in a classroom scenario?
3. What effect do you hope your favorite will have?

Fig. 4. The left image provides an overview of the high school's library. The right image shows the setup of the study prior to the start of an experimental trial.

4.2 Procedure

We conducted the study in the library of a German high school as displayed in Fig. 4. There were always 2–3 students participating in the study at the same time. After welcoming the participants, we briefly explained the experiment. The students randomly chose one of the three conditions by drawing a labeled card. They were then assigned to workstations where they were first to read the student participant information. Once all students completed this step, they completed the pre-trial phase consisting of the demography questionnaire, PANAS-C, and pre-trail exam. Subsequently, they started the learning phase with the respective version of Horst. To ensure that each student knew how to use their learning environment, we provided a separate learning instruction for each condition.

preprint

The learning time was limited. All students in the *Horst-WS* condition had 15 minutes and all students in the *Horst-S* and *Horst-TAR* condition had 14 minutes. We shortened the learning time for the two digital versions as their tutorials already explain the first organ and hence the participants start the learning phase while familiarizing them with the learning environment. After the learning phase, the experimenters collected the learning materials and handed out the post-trial materials. These materials consisted of Paas scale, PANAS-C, KIM, post-trial exam, and feedback questions. The sequence of the materials ensured a direct assessment of the cognitive load, current affect and motivation, while reducing the chance for a recency effect when completing the exam.

4.3 Participants

The sample was recruited from the pupils attending the sixth-grade at a German highschool. All parents received an information sheet in advance describing the purpose and procedure of the study, COVID-19 infection control measures, voluntariness and anonymity, data protection and handling of the anonymized data. Based on this, they were able to make a decision as to whether they agreed to their child's participation in the study. Only children who had a signed parental consent form and voluntarily wished to partake in the study participated.

A total of $N = 68$ high school students participated in the study. However, the data sets of 4 students had to be excluded from the analysis because of missing information and conspicuous answers in the quantitative and qualitative data. In the end, the data of $N = 64$ students were evaluated. Of these, 27 subjects were male (42.86%), 36 subjects were female (56.25%), and 1 subject was diverse (1.56%). The average age was 11.98 years ($SD = 0.42$). The majority of students (56; 87.50%) reported speaking German at home. No student showed comprehension problems during the study. The majority of students had quite a bit or a lot (27 each; 42.19%) of experience with smartphone apps at the time of the survey, while a small proportion had little (9; 14.06%) or no experience (1; 1.56%) with them. In terms of experience with AR apps, the ratio was reversed. Thus, the majority had no (17; 26.56%) or little (24; 37.50%) experience with AR apps and a minority had quite a bit (16; 25.00%) or a lot (7; 10.94%) of experience. Interest in biology was in the middle range on the scale of 1 "Not at all" to 4 "Very much" ($M = 2.45$; $SD = 0.66$).

5 Results

We tested the reliability of the scales by computing Cronbach's α if applicable. To compare the different learning environments, we calculated one-factor analyses of variance (ANOVA) for the results of the PANAS-C, Paas scale, KIM, and score on the post exam. As a measure of effect size, we computed the partial eta squared ${\eta_p}^2$. Homoscedasticity was checked using Levene's test before each ANOVA. Alternatively, if there was no homogeneity of variances, a Welch

preprint

Table 1. Descriptive statistics; $N = 64$. Values are $M(SD)$.

Scale	**Horst-WS** ($n = 22$)	**Horst-S** ($n = 22$)	**Horst-TAR** ($n = 20$)
Exam			
Pre	2.55 (2.13)	2.91 (2.27)	3.10 (2.40)
Post	12.32 (4.08)	11.98 (5.07)	10.25 (5.28)
PANAS-C			
PA_{pre} ($\alpha = .75$)	3.15 (0.92)	3.14 (0.65)	3.23 (0.61)
PA_{post} ($\alpha = .90$)	2.62 (1.08)	2.99 (0.91)	3.15 (0.69)
NA_{pre} ($\alpha = .58$)	1.31 (0.29)	1.28 (0.38)	1.36 (0.50)
NA_{post} ($\alpha = .47$)	1.25 (0.28)	1.14 (0.19)	1.24 (0.35)
KIM			
Enjoyment ($\alpha = .91$)	2.86 (1.25)	3.95 (1.00)	3.82 (0.95)
Competence ($\alpha = .81$)	3.15 (0.91)	3.68 (0.83)	3.60 (0.58)
Choice ($\alpha = .85$)	3.84 (1.19)	3.86 (0.97)	4.00 (0.99)
Pressure ($\alpha = .71$)	2.83 (1.08)	2.08 (0.70)	2.23 (0.91)
Paas			
Total score	5.55 (1.68)	4.59 (1.68)	5.25 (1.68)
Feedback			
Repetition of learning	3.32 (1.55)	4.05 (1.05)	3.25 (1.37)
Technology for other subjects	2.50 (1.44)	4.32 (0.95)	3.65 (1.18)
Grade	2.86 (1.24)	1.91 (1.15)	1.93 (0.65)

ANOVA was calculated. Since ANOVA is robust to violation of the normal distribution, no corrective action was taken in case of violation. Some analyses of variance, including PA_{post}, NA_{post}, and post exam, required consideration of one or more covariates (ANCOVA). In this case, an additional ANOVA was computed to ensure the independence of the covariates and the learning environments. If an omnibus test produced a significant difference, we calculated Tukey-Kramer comparisons for unequal sample sizes or Games-Howell tests for unequal variances and unequal sample sizes. The descriptive statistics are displayed in Table 1.

5.1 Learning Effectiveness

We analyzed the knowledge gain of the participants by computing two-sided repeated measures t-tests. All learning environments caused a significant knowledge gain with a strong effect size, *Horst-WS* $t(21) = 9.86$, $p < .001$, $d = 2.10$; *Horst-S* $t(21) = 9.94$, $p < .001$, $d = 2.12$; *Horst-TAR* $t(19) = 7.51$, $p < .001$, $d = 1.68$.

Computing an ANCOVA revealed that the $exam_{pre}$ score was significantly related to the $exam_{post}$ score, $F(1, 60) = 12.27$, $p < .001$, ${\eta_p}^2 = .17$. However, the $exam_{post}$ score did not differ significantly between conditions, $F(2, 60) = 1.84$, $p = .17$, ${\eta_p}^2 = .06$.

5.2 Emotion

Computing an ANCOVA of PA_{post} with PA_{pre} as a covariate showed that PA_{pre} was significantly related to PA_{post}, $F(1, 60) = 76.16$, $p < .001$, ${\eta_p}^2 = .56$. Moreover, PA_{post} differed significantly between the conditions with a medium effect size controlling for PA_{pre}, $F(2, 60) = 3.60$, $p = .03$, ${\eta_p}^2 = 0.11$. Post-hoc Tukey-Kramer comparisons substantiated that *PA* was significantly higher after using *Horst-TAR* than after using *Horst-WS*, $t(60) = 2.46$, $p = .04$. The differences of *Horst-WS* and *Horst-S*, $t(60) = 2.14$, $p = .09$, and *Horst-S* and *Horst-TAR*, $t(60) = .37$, $p = .97$, were not significant.

Computing an ANCOVA of NA_{post} with NA_{pre} as a covariate showed a significant relationship of NA_{post} and NA_{pre}, $F(1, 60) = 48.83$, $p < .001$, ${\eta_p}^2 = .45$. We found no significant differences between the learning environments with respect to *NA* controlling for NA_{pre}, $F(2, 60) = 1.26$, $p = .29$, ${\eta_p}^2 = .04$.

5.3 Motivation

The analyses of the subscales *choice*, $F(2, 61) = .14$, $p = .87$, ${\eta_p}^2 = .01$, and *competence*, $F(2, 61) = 2.82$, $p = .07$, ${\eta_p}^2 = .09$, did not reveal any statistically significant differences. In contrast, significant effects with a medium effect size were detected for the *pressure* scale, $F(2, 61) = 4.21$, $p = .02$, ${\eta_p}^2 = .12$. Subsequent Tukey-Kramer comparisons showed that subjects in the *Horst-S* condition felt significantly less pressure than subjects in the *Horst-WS* condition, $t(61) = -2.76$, $p = .02$. *Horst-TAR* and *Horst-WS*, $t(61) = -2.13$, $p = .09$, and *Horst-S* and *Horst-TAR*, $t(61) = .56$, $p = .84$, did not differ significantly. Also, we found a significant difference with a strong effect size for the *enjoyment* subscale, $F(2, 61) = 6.57$, $p = .003$, ${\eta_p}^2 = .18$. Post-hoc comparisons revealed that *Horst-S*, $t(61) = 3.34$, $p = .004$, and *Horst-TAR*, $t(61) = 2.86$, $p = .02$, caused higher learner interest and enjoyment than *Horst-WS*. *Horst-S* and *Horst-TAR* were not significantly different, $t(61) = -.39$, $p = .92$.

These effects remained stable when controlling for the possible covariates *experience with AR* and *interest* by testing with an ANCOVA, independence of covariates given, $p > .05$; $F(2, 59) = 7.10$, $p = .002$, ${\eta_p}^2 = .19$.

5.4 Cognitive Load

A one-factor ANOVA revealed no significant differences between the conditions with respect to the *Paas* total score, $F(2, 61) = 1.85$, $p = .17$, ${\eta_p}^2 = .06$.

5.5 Feedback

We found no significant differences with regard to the willingness to repeat learning with the previously tested learning environment, $F(2, 61) = 2.34$, $p = .11$, ${\eta_p}^2 = .07$.

Calculating a Welch ANOVA, we found significant differences with a strong effect size between the conditions for the *desire to use the same technology for*

preprint

learning different subjects, $F(2, 39) = 12.13$, $p < .001$, ${\eta_p}^2 = .38$. Subsequent Games-Howell tests showed that the desire was significantly higher for *Horst-S*, $t(36.3) = 4.95$, $p < .001$, and *Horst-TAR*, $t(39.6) = 2.84$, $p = .02$ than for *Horst-WS*. We did not find a significant difference between *Horst-S* and *Horst-TAR*, $t(36.4) = 2.01$, $p = .12$.

Computing an ANOVA revealed a significant difference with a strong effect size for the *grades of the learning environments*, $F(2, 61) = 5.80$, $p = .005$, ${\eta_p}^2 = .16$. Post-hoc Tukey-Kramer comparisons showed that *Horst-S*, $t(61) = -3.00$, $p = .01$, and *Horst-TAR*, $t(61) = -2.88$, $p = .02$, were graded significantly better than *Horst-WS*. We did not find a significant difference between *Horst-S* and *Horst-TAR*, $t(61) = .05$, $p = .99$.

5.6 Teacher Interview

Two teachers preferred *Horst-S*, while one teacher preferred *Horst-TAR*. They justified the preference for *Horst-S* with the self-explanatory and intuitive design as well as the faster completion time compared to *Horst-TAR*. With respect to *Horst-TAR*, the teachers stated that the technology is novel and represents a compromise between digital teaching via app and classical biology teaching with props.

Different approaches were listed for the possible integration of the learning environments into classroom teaching. *Horst-S* could be used to repeat the most recently acquired knowledge at the beginning of a lesson. For instance, students could use the free dissection mode to look up information on individual organs. Furthermore, *Horst-S* and *Horst-TAR* could be used to enhance a lesson. Using this approach, small groups of up to four students could complete the assisted dissection. However, it would be important to contextualize the learning content beforehand and, if necessary, to provide an introduction to the respective learning environment. For example, the teachers mentioned the body structure of the frog in comparison to other vertebrates, such as dogs, cats, pigs, and horses, as a possible context. They also emphasized that students should be taught additional information about the usual process of dissection in contrast to digital dissection. It would be of great importance for students to recognize that a real dissection involves the death of an animal and is distinctly different olfactory, visually, and tactilely. Finally, it is also necessary to discuss and consolidate the knowledge acquired after finishing the simulation.

The teachers expected a gain of knowledge and the development of digital competencies among students when using *Horst-S*. By using smartphones in classroom, students should no longer regard them only as a means of entertainment, but also as an easy access to knowledge. For *Horst-TAR*, a greater interest, sustainability and transfer of the acquired knowledge was assumed. The teachers expected using this learning environment to be fun for students.

preprint

6 Discussion

The present study investigated the effects of TAR technology on effectiveness, emotion, motivation, and cognitive load in an educational context.

6.1 Learning Effectiveness

All learning environments successfully led to a significant knowledge gain of the participants over the course of the learning period. As learning is a multidimensional concept that has proven challenging to measure in previous studies [37] and rarely resulted in statistical differences in comparative media studies [53], the lack of a statistical difference is not surprising. As displayed in Table 1, the participants started with a very low knowledge level before the experiment and yielded a significant improvement in their exam score after the learning phase. Although we did not find a significant difference between the conditions, this result is promising. It supports that digital learning environment are comparable in their learning effectiveness to established approaches, but cause a stronger motivation and higher enjoyment during the learning process [38]. Overall, this validates the approach of technology-based learning and can be a notable insight for teachers searching for new learning methods. Despite these promising results, we need to reject **H1**.

6.2 Emotion

The analysis of the NA scales show that NA_{post} did not differ significantly between the conditions. The negative affect declined over the learning phase and generally was at a low level. This indicates that no condition evoked negative emotions in the participants during the learning phase. PA_{post} also declined over the learning period, but remained in the middle range. In direct comparison, participants of the *Horst-TAR* condition reported a significantly higher PA_{post} than the participants of the *Horst-WS* condition. This suggests that the interactions and higher user experience of the TAR learning environment evoked more positive emotions in the participants. Although not significantly different, *Horst-S* also yielded a higher PA_{post} score than the traditional worksheet approach. Thus, we need to reject **H2**. Although not fully supporting our hypothesis, our results support the benefits of using digital and gamified learning environments over traditional learning methods for achieving an emotionally positive learning process.

6.3 Motivation

The analysis of the KIM subscales revealed no significant differences for *choice* and *competence*. However, we found significant differences in the results of the *pressure* subscale. Participants felt moderately tense while learning with *Horst-WS*, whereas participants were more relaxed during the learning phase with

preprint

Horst-S and *Horst-TAR*. Surprisingly, *Horst-S* yielded the lowest score on the pressure subscale. This effect could be explained by the different interactions with the two learning environments. The smartphone version only requires touch interactions, whereas the TAR version requires users to extract and manipulate physical objects in addition to holding the smartphone. Similarly, the participants reported a significantly higher *enjoyment and interest* after using *Horst-S* and *Horst-TAR* than *Horst-WS*. This result supports our PANAS-C measurements indicating a higher positive effect. Our measurements provide further evidence of the benefits of using smartphones and gamification for educational purposes [34].

Taken together, we can assume that the two digital learning environments cause a higher motivation during learning which leads to a better learning experience. However, we need to reject **H3** as TAR did not evoke a higher motivation than the smartphone-only version.

6.4 Cognitive Load

We did not find any significant differences in the Paas measurements between the three conditions. Assessing the cognitive load of the overall learning process, this result is still of high importance. It indicates that the two digital learning environments invoked no cognitive overload in the participants compared to the traditional method. Our results support that high user experience and the application of multimedia principles [35], such as coherence, signaling, segmentation, personalization, embodiment, and spatial and temporal proximity principles, prove to be reliable methods to avoid overloading cognitive resources. This suggests that smartphone and TAR learning environments following these design principles can safely be integrated in classroom teaching without risking to cause negative side effects. However, we need to reject **H4**.

6.5 Feedback

The willingness to repeat the learning with *Horst-WS* and *Horst-TAR* was in the middle range. *Horst-S* yielded an above average score. While not significantly different, the scores support the positive results of the measurements of emotion and motivation. The digital learning environments led to an overall better learning experience. This outcome is supported by the significant differences in the desire to use the same technology for learning different subjects. In particular, *Horst-S* and *Horst-TAR* evoked a significantly higher desire compared to *Horst-WS*. The participants saw benefits in using technology-based learning. Finally, the participants graded *Horst-S* and *Horst-TAR* significantly better than *Horst-WS*. This feedback supports the overall results of our study.

6.6 Implications

We found no significant difference between *Horst-TAR* and *Horst-S* for any of the measured qualities. Our measurements suggest a higher positive emotion when

preprint

Table 2. Comparison of worksheet, smartphone, and TAR learning environments.

	Worksheet	**Smartphone**	**TAR**
Material	Highest accessibility	Moderate accessibility	Lowest accessibility
	Paper	Smartphone	Smartphone & tangible markers
Features	Lowest time expenditure	Moderate time expenditure	Highest time expenditure
	No operation instruction needed	Operation instruction required	Operation instruction required
	Cannot display 3D	Can display 3D & animations	Can display 3D & animations in a natural way
Positive aspects	Efficient & effective, medium cognitive load	Effective, better affect, less pressure, higher enjoyment, medium cognitive load, high appreciation	Effective, higher enjoyment, medium cognitive load, high appreciation

using *Horst-TAR* than *Horst-WS*. In contrast, *Horst-S* outperformed the traditional learning method with respect to perceived pressure. Both digital learning environments evoked a higher enjoyment during the learning process. Thus, our study suggests that both tested technologies are equally beneficial for learning compared to the traditional worksheet method. Gamified smartphone and TAR learning environments indicate to evoke a better learning experience while yielding a similar learning effectiveness to traditional approaches. A different study already showed a significantly higher stimulation and attractiveness of *Horst-TAR* compared to its AR and smartphone counterparts [40]. In combination with our insights, TAR indicates the potential to cause a better overall learning experience. This, however, comes at the price of a higher preparation complexity for classroom integration.

Based on the assessment of three teachers, it is important to contextualize the learning environments when embedding them in a lesson and to reflect on the learning content after the learning phase. As using TAR learning environments also requires teachers to provide augmented objects for each learner, this positive aspect should ideally be used for the introduction of very complex learning content to add an additional stimulating aspect to the learning process. For all other subjects, smartphone learning environments might provide the best alternative to traditional worksheet approaches. They can be self-explanatory, intuitive, and easy to use. Additionally, using smartphones enables students to develop important digital competencies. Table 2 provides a direct comparison of the positive effects of using a worksheet, smartphone, and TAR learning environments.

preprint

6.7 Limitations

Our results could have been influenced by a novelty effect. Only 23 participants reported a high experience with AR apps. In addition, the integration of smartphones is not common praxis, yet. This could have resulted in a higher motivation and general more positive feedback when a digital learning environment was used. However, our statistical analysis revealed no effect of previous experience with AR on the reported motivation.

In contrast to *Horst-WS*, both digital versions included gamification. This might have confounded the evaluation of the experienced emotions and motivation. However, this influence is only true for direct comparisons to *Horst-WS*. Both digital versions used the same gamification elements and hence the comparison of emotions and motivation should not be confounded in a direct comparison between them.

Finally, the subjective feedback of the participants on their condition could have been influenced by an effect of social desirability. This might have confounded the results of the feedback questions.

7 Conclusion

In the present study, we investigated the effects of different technologies on the learning effectiveness, emotion, motivation, and cognitive load of a biology learning environment. Our user study at a local high school found no significant difference between conditions in learning effectiveness and cognitive load. However, the TAR version yielded a significantly higher positive emotion in comparison to the worksheet conditions while not being significantly different to the smartphone version. The two digital learning environments evoked a significantly higher motivation in comparison to the traditional worksheet approach. These findings can be of high importance for teachers and developers. While TAR can cause higher positive emotions, it is also more complex with respect to the required materials. Thus, TAR potentially is more suited for the introduction of new and very challenging learning content to benefit from the positive emotions, whereas smartphone learning environments are generally more practical for a very motivated learning process. Future work shall investigate whether our results are affected by a novelty effect. Also, a long-term study shall investigate whether the tested technologies improve the retention of learning content. Finally, research shall evaluate whether true three-dimensional targets, such as 3D-printed organ markers, can improve the learning process even further.

Acknowledgements We would like to thank principal Wolfgang Naumann and the teachers of LuO Darmstadt Gymnasium for their support of our study.

This research was performed within the *"Die Zukunft des MINT-Lernens"* project, supported by the Deutsche Telekom Stiftung.

preprint

References

1. Antle, A., Wise, A.: Getting down to details: Using learning theory to inform tangibles research and design for children. Interacting with Computers **25**, 1–20 (01 2013). https://doi.org/10.1093/iwc/iws007
2. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators and Virtual Environments **6**(4), 355–385 (1997)
3. Baack, K., Göbel, R., Maier, A., Marx, U., Remé, R., Seitz, H.J.: Natura 1. Biologie für Gymnasien. Klett (2013)
4. Bacca, J., Baldiris, S., Fabregat, S., Graf, S., Kinishuk: Augmented reality trends in education: A systematic review of research and applications. Educational Technology & Society **17**(4), 133–149 (2014)
5. Bach, B., Sicat, R., Beyer, J., Cordeil, M., Pfister, H.: The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? IEEE Transactions on Visualization and Computer Graphics **24**(1), 457–467 (2018)
6. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and magic lenses: The see-through interface. In: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques. pp. 73–80. SIGGRAPH '93, Association for Computing Machinery, New York, NY, USA (1993). https://doi.org/10.1145/166117.166126
7. für Schulqualität und Bildungsforschung München, S.: Lehrplanplus. available at https://www.lehrplanplus.bayern.de/ (2020)
8. Billinghurst, M., Kato, H., Poupyrev, I.: Tangible augmented reality. ACM SIGGRAPH ASIA 2008 Courses (01 2008). https://doi.org/10.1145/1508044.1508051
9. Breyer, B., Bluemke, M.: Deutsche version der positive and negative affect schedule panas (gesis panel) (03 2016). https://doi.org/10.6102/zis242
10. Cabero-Almenara, J., Roig-Vila, R.: The motivation of technological scenarios in augmented reality (ar): Results of different experiments. Applied Sciences **9**(14) (2019). https://doi.org/10.3390/app9142907
11. Carrera, C.C., Perez, J.L.S., de la Torre Cantero, J.: Teaching with ar as a tool for relief visualization: usability and motivation study. International Research in Geographical and Environmental Education **27**(1), 69–84 (2018). https://doi.org/10.1080/10382046.2017.1285135
12. Chandrasekera, T., Yoon, S.: The effect of tangible user interfaces on cognitive load in the creative design process. In: 2015 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design. pp. 6–8 (2015). https://doi.org/10.1109/ISMAR-MASHD.2015.18
13. Dengel, A., Mägdefrau, J.: Immersive learning explored: Subjective and objective factors influencing learning outcomes in immersive educational virtual environments. In: 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). pp. 608–615 (2018). https://doi.org/10.1109/TALE.2018.8615281
14. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11). pp. 9–15. ACM, Tampere, Finland (Sep 2011). https://doi.org/10.1145/2181037.2181040
15. Diegmann, P., Schmidt-Kraepelin, M., Van den Eynden, S., Basten, D.: Benefits of augmented reality in educational environments - a systematic literature review.

preprint

In: Proceedings of the 12th International Conference on Wirtschaftsinformatik. pp. 1542–1556. Osnabrück, Germany (March 2015)
16. Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., Chorpita, B.: The 10-item positive and negative affect schedule for children, child and parent shortened versions: Application of item response theory for more efficient assessment. Journal of Psychopathology and Behavioral Assessment **34** (06 2012). https://doi.org/10.1007/s10862-011-9273-2
17. de Freitas, S., Liarokapis, F.: Serious games: A new paradigm for education? In: Ma, M., Oikonomou, A., Jain, L.C. (eds.) Serious Games and Edutainment Applications, pp. 9–23. Springer-Verlag, London (2011). https://doi.org/10.1007/978-1-4471-2161-9_2
18. Garzón, J., Pavón, J., Baldiris, S.: Systematic review and meta-analysis of augmented reality in educational settings. Virtual Reality **23**, 447–459 (2019). https://doi.org/10.1007/s10055-019-00379-9
19. Gervautz, M., Schmalstieg, D.: Anywhere interfaces using handheld augmented reality. Computer **45**(7), 26–31 (2012). https://doi.org/10.1109/MC.2012.72
20. Gikas, J., Grant, M.M.: Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. The Internet and Higher Education **19**, 18–26 (2013). https://doi.org/10.1016/j.iheduc.2013.06.002
21. Glandt, D.: Amphibien und Reptilien. Herpetologie für Einsteiger. Springer Spektrum (2016), https://doi.org/10.1007/978-3-662-49727-2
22. Gutiérrez Posada, J.E., Hayashi, E.C.S., Baranauskas, M.C.C.: On feelings of comfort, motivation and joy that gui and tui evoke. In: Marcus, A. (ed.) Design, User Experience, and Usability. User Experience Design Practice. pp. 273–284. Springer International Publishing, Cham (2014)
23. Ishii, H., Ullmer, B.: Tangible bits: Towards seamless interfaces between people, bits and atoms. In: Proceedings of the 1997 CHI Conference on Human Factors in Computing Systems (CHI '97). pp. 234–241. Atlanta, USA (1997). https://doi.org/10.1145/258549.258715
24. Kalemkuş, J., Kalemkuş, F.: Effect of the use of augmented reality applications on academic achievement of student in science education: meta analysis review. Interactive Learning Environments (2022). https://doi.org/10.1080/10494820.2022.2027458
25. Keller, S., Rumann, S., Habig, S.: Cognitive load implications for augmented reality supported chemistry learning. Information **12**(3), 96–115 (2021). https://doi.org/10.3390/info12030096
26. Khan, M., Trujano, F., Choudhury, A., Maes, P.: Mathland: Playful mathematical learning in mixed reality. In: CHI'18 Extended Abstracts. Montréal, Canada (April 2018)
27. Kim, K., Billinghurst, M., Bruder, G., Duh, H.B., Welch, G.F.: Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). IEEE Transactions on Visualization and Computer Graphics **24**(11), 2947–2962 (2018). https://doi.org/10.1109/TVCG.2018.2868591
28. Koehler, M., Mishra, P.: What happens when teachers design educational technology? the development of technological pedagogical content knowledge. Journal of Educational Computing Research **32**(2), 131–152 (2005)
29. Küçük, S., Kapakin, S., Göktaş, Y.: Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. Anatomical Sciences Education **9**(5), 411–421 (2016). https://doi.org/10.1002/ase.1603

preprint

30. Laurent, J., Catanzaro, S.J., Joiner, T.E., Rudolph, K.D., Potter, K.I., Lambert, S.F., Osborne, L.N., Gathright, T.: A measure of positive and negative affect for children: Scale development and preliminary validation. Psychological Assessment **11**, 326–338 (1999)
31. Li, J., Van der Spek, E.D., Feijs, L., Wang, F., Hu, J.: Augmented reality games for learning: A literature review. In: International Conference on Distributed, Ambient, and Pervasive Interactions (2017). https://doi.org/10.1007/978-3-319-58697-7_46
32. Liu, L., Wagner, C., Suh, A.: Understanding the success of pokémon go: Impact of immersion on players' continuance intention. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments. pp. 514–523. Springer International Publishing, Cham (2017)
33. Loderer, K., Pekrun, R., Lester, J.C.: Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. Learning and Instruction **70** (2018)
34. Majuri, J., Koivisto, J., Hamari, J.: Gamification of education and learning: A review of empirical literature. In: Proceedings of the 2nd international GamiFIN conference (GamiFIN 2018). pp. 11–19 (2018)
35. Mayer, R.E.: Cognitive theory of multimedia learning. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning, pp. 43–71. Cambridge University Press (2014). https://doi.org/10.1017/CBO9781139547369.005
36. McGonigal, J.: Reality is Broken: Why Games Make Us Better and How They Can Change the World. Penguin Press, New York, 1st edn. (2011)
37. Ni, A.Y.: Comparing the effectiveness of classroom and online learning: Teaching research methods. Journal of Public Affairs Education **19**, 199–215 (06 2013). https://doi.org/10.1080/15236803.2013.12001730
38. Oberdörfer, S.: Better Learning with Gaming: Knowledge Encoding and Knowledge Learning Using Gamification. Ph.D. thesis, University of Würzburg (2021). https://doi.org/10.25972/OPUS-21970
39. Oberdörfer, S., Birnstiel, S., Latoschik, M.E., Grafe, S.: Mutual benefits: Interdisciplinary education of pre-service teachers and hci students in vr/ar learning environment design. Frontiers in Education **6**, 233 (2021). https://doi.org/10.3389/feduc.2021.693012
40. Oberdörfer, S., Elsässer, A., Grafe, S., Latoschik, M.E.: Grab the frog: Comparing intuitive use and user experience of a smartphone-only, ar-only, and tangible ar learning environment. In: Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21). Toulouse & Virtual, France (September 2021). https://doi.org/10.1145/3447526.3472016
41. Oberdörfer, S., Elsässer, A., Schraudt, D., Grafe, S., Latoschik, M.E.: Horst – the teaching frog: Learning the anatomy of a frog using tangible ar. In: Proceedings of the 2020 Mensch und Computer Conference (MuC '20). pp. 303–307. Magdeburg, Germany (September 2020). https://doi.org/10.1145/3404983.3410007
42. Oberdörfer, S., Latoschik, M.E.: Gamified knowledge encoding: Knowledge training using game mechanics. In: Proceedings of the 10th International Conference on Virtual Worlds and Games for Serious Applications (VS Games '18). ©2018 IEEE. Reprinted, with permission., Würzburg, Germany (September 2018). https://doi.org/10.1109/VS-Games.2018.8493425
43. Paas, F.G.W.C.: Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology **84**(4), 429–434 (1992)

preprint

44. Paas, F.G.W.C., Ayres, P., Pachman, M.: Assessment of cognitive load in multimedia learning - theory, methods and applications. In: Robinson, D.H., Schraw, G. (eds.) Recent Innovations in Educational Technology that Facilitate Student Learning, pp. 11–35. Information Age Publishing Inc. (2008)
45. Pekrun, R.: The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educational Psychology Review **18**, 315–341 (12 2006). https://doi.org/10.1007/s10648-006-9029-9
46. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary Educational Psychology **61** (2020). https://doi.org/10.1016/j.cedpsych.2020.101860
47. Ryan, R., Deci, E.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. The American Psychologist **55**, 68–78 (February 2000). https://doi.org/10.1037/0003-066X.55.1.68
48. Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies **74**, 14–31 (2015). https://doi.org/10.1016/j.ijhcs.2014.09.006
49. Steinhaeusser, S.C., Riedmann, A., Haller, M., Oberdörfer, S., Bucher, K., Latoschik, M.E.: Fancy fruits - an augmented reality application for special needs education. In: Proceedings of the 11th International Conference on Virtual Worlds and Games for Serious Applications (VS Games '19). IEEE, Vienna, Austria (September 2019)
50. Storch, V., Welsch, U.: Amphibia, Lurche. Springer Spektrum (2014). https://doi.org/10.1007/978-3-642-41937-9_13
51. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Sciences **12**(2), 257–285 (1988)
52. Sweller, J.: Measuring cognitive load. Perspectives on medical education **7**(1), 1–2 (2018)
53. Tulodziecki, G., Herzig, B., Grafe, S.: Medienbildung in Schule und Unterricht. 3. vollst. akt. u. überarb. Auflage. Klinkhardt, Bad Heilbrunn (2021)
54. Weerasinghe, M., Biener, V., Grubert, J., Quigley, A.J., Toniolo, A., Pucihar, K.Č., Kljun, M.: Vocabulary: Learning vocabulary in ar supported by keyword visualisations. arXiv, 2207.00896, cs.HC (2022). https://doi.org/10.48550/ARXIV.2207.00896
55. Wilde, M., Bätz, K., Kovaleva, A., Urhahne, D.: Überprüfung einer kurzskala intrinsischer motivation (kim) (2009)
56. Yin, X., Li, G., Deng, X., Luo, H.: Enhancing k-16 science education with augmented reality: A systematic review of literature from 2001 to 2020. In: 2022 8th International Conference of the Immersive Learning Research Network (iLRN). pp. 215–219 (2022). https://doi.org/10.23919/iLRN55037.2022.9815958
57. Youngblut, C.: Use of multimedia technology to provide solutions to existing curriculum problems: Virtual frog dissection. Ph.D. thesis, George Mason University, United States (2001)
58. Zuckerman, O., Gal-Oz, A.: To tui or not to tui: Evaluating performance and preference in tangible vs. graphical user interfaces. International Journal of Human-Computer Studies **71**(7–8), 803–820 (2013). https://doi.org/10.1016/j.ijhcs.2013.04.003