preprint

Grab the Frog: Comparing Intuitive Use and User Experience of a Smartphone-only, AR-only, and Tangible AR Learning Environment

SEBASTIAN OBERDÖRFER, University of Würzburg, Germany
ANNE ELSÄSSER, University of Würzburg, Germany
SILKE GRAFE, University of Würzburg, Germany
MARC ERICH LATOSCHIK, University of Würzburg, Germany

Fig. 1. We compare the perceived intuitive use and user experience of a learning environment developed for smartphone (left), handheld AR (middle), and tangible AR (right).

The integration of Augmented Reality (AR) in teaching concepts allows for the visualization of complex learning contents and can simultaneously enhance the learning motivation. By providing Tangible Augmented Reality (TAR), an AR learning environment receives a haptic aspect and allows for a direct manipulation of augmented learning materials. However, manipulating tangible objects while using handheld AR might reduce the intuitive use and hence user experience. Users need to simultaneously control the application and manipulate the tangible object. Therefore, we compare the differences in intuitive use and user experience evoked by varied technologies of knowledge presentation in an educational context. In particular, we compare a TAR learning environment targeting the learning of the anatomy of vertebrates to its smartphone-only and AR-only versions. The three versions of the learning environment only differ in their method of knowledge presentation. The three versions show a similar perceived intuitive use. The TAR version, however, yielded a significantly higher attractiveness and stimulation than AR-only and smartphone-only. This suggests a positive effect of TAR learning environments on the overall learning experience.

CCS Concepts: • **Applied computing → Education**; • **Human-centered computing**;

Additional Key Words and Phrases: augmented reality; education; serious games; gamification; tangible user interfaces

ACM Reference Format:
Sebastian Oberdörfer, Anne Elsässer, Silke Grafe, and Marc Erich Latoschik. 2021. 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), September 27-October 1, 2021, Toulouse & Virtual, France.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3447526.3472016

1 INTRODUCTION

Learning new knowledge is a challenging task that requires a high degree of motivation, discipline, and high amount of work [49]. The widespread availability of mobile devices, especially smartphones, among students allows for technology-based teaching concepts [20]. Besides providing access to smartphone-only learning environments, smartphones further enable an easy integration of Augmented Reality (AR) in classrooms [3]. Combining physical objects belonging to the learning content with so-called fiducial markers achieves *Tangible AR* (*TAR*) [7].

Using AR for educational purposes can lead to an increased motivation, a gain of experience due to the direct application of the learning contents, and higher task performance [3]. In combination with tangible user interfaces, the learning process further receives a physical aspect [7]. TAR provides six degrees of freedom, thus allowing learners to inspect augmented objects from all angles by manipulating them directly [4]. This not only enables learners to develop a spatial understanding for the learning content, but also intensifies the direct application of the knowledge. The motivating aspects can further be improved when following a gamified approach [38, 40].

However, TAR using smartphones may result in complex interactions potentially negatively affecting the intuitive use and user experience. Learners not only need to interact with the application running on the smartphone, but also manipulate a tangible object at the same time. This could reduce the overall effectiveness of handheld TAR. Therefore, it is crucial to compare TAR learning environments to state-of-the-art smartphone-only and AR-only applications with regard to intuitive use and user experience.

Contribution

In the present study, we compare the TAR application *Horst – The Teaching Frog* [39] dealing with the learning of the anatomy of vertebrates to its smartphone-only and AR-only versions (see Figure 1). Using the application, learners can virtually dissect a frog, thus acquiring knowledge about its anatomy and the functions of the organs. The three versions of the learning environment only differ with regard to the knowledge presentation method, i.e., smartphone-only, AR-only, and TAR. In a user study, the three versions show a similar perceived intuitive use. The applications, however, differ with regard to user experience. The TAR version yielded a significantly higher attractiveness and stimulation in contrast to the other two versions. As TAR applications can elicit the same intuitive use, their higher user experience suggests a higher positive effect on the overall learning experience.

2 THEORETICAL BACKGROUND

AR three-dimensionally integrates virtual elements into the real-world that are interactive in real time [2]. Using AR applications in an educational context can lead to an increased motivation, improvements with respect to interaction and collaboration, and gaining experience with the direct application of the learning contents [3]. Also, AR learning environments provide a spatial and direct visualization of the learning contents [13].

2.1 Using AR for Educational Purposes

Although AR learning applications target almost any kind of area, they mostly focus STEM related topics [3]. For instance, *Mathland* demonstrates the mathematics behind the Newtonian physics and allows users to modify and thus explore the physical laws [26]. The direct interaction with real-world objects can further support the requirements of special needs education [45].

Users can experience AR using headworn, handheld, and projected displays [27]. While headworn and projected displays provide a greater freedom to users, handheld AR is especially suited for learning environments due to the widespread availability of smartphones. Handheld AR turns the smartphone into a "Magic Lense" [6], thus revealing the augmentations. Using smartphone AR also has the advantage of a high familiarity of the users with the device [33].

AR elements can be manipulated using a tangible user interface [19]. A tangible user interface connects digital information with real world objects, i.e., using the objects as input and output devices [25]. A TAR interface registers virtual information to real-world objects, thus allowing for their augmentation and an interaction by manipulating the physical objects [7]. This approach renders an AR system very intuitive [7] and especially suits the visualization of 3D models [4]. With respect to learning, TAR can reduce cognitive load, intensify work on learning material, improve usability, support mental skills and collaboration [1]. Tangible user interfaces further evoke a higher degree of joy as well as motivation in comparison to graphical user interfaces [21] and reduce the cognitive load [9]. This could lead to a higher learning motivation and learning effectiveness of TAR learning environments.

2.2 Enhancing the Learning Processes

Developers can use the powerful graphics engines of current game engines, e.g., *Unreal Engine* [24] and *Unity* [47]. These game engines can visualize complex information and are even used to demonstrate complex scientific problems [28, 29, 35]. This capability can even become more prominent when registering the virtual information to the real-world. In this case, learners can visualize complex problems using an AR learning environment, thus potentially developing a better understanding for them.

The overall learning process can become an engaging, vivid, and inspiring experience when targeting a gamified approach [36]. Gamified learning environments can either be *serious games* [12] or non-gaming learning applications enhanced by *gamification* [44]. These learning environments combine classical game design elements with pedagogical aspects to yield a high learning motivation [10]. Mapping the learning contents to central game mechanics or core interactions achieves their application and demonstration during the simulation [42].

2.3 Usability Design Frameworks

Research proposes various design frameworks, design guidelines, and heuristics to maximize the usability of AR and TAR applications [15, 23, 30, 32]. For instance, since smartphone displays are rather small, the amount of shown information should be limited to a minimum to reduce the cognitive load of a user [23]. Also, controlling the application should be possible with a few interactions only, e.g., using buttons and menus that appear at a fixed position [23]. For ensuring a good usability, the visualization of icons, font size, size of interactive elements, user guidance, and single-handed controls should match the technology used, e.g., be adjusted for smartphone displays [32]. The single-handed controls are especially of high relevance when targeting TAR. Here, users need to simultaneously use a handheld device while manipulating physical objects. This could lead to complex interactions, thus negatively affecting the intuitive use and hence the learning effectiveness. To minimize this effect, not only the handheld device, but also TAR objects need to allow for single-handed manipulation. Overall, this requires TAR applications to be designed for two-handed interactions. Following this design goal could lead to a high intuitive use despite complex interactions. The design of the TAR objects should further exploit the 3D space for rich interactions and provide continuous as well as seamless interactions [22]. Hence, the TAR application should continuously track tangible objects, instantaneously react to changes, and support complex interactions, e.g., disassembling an object.

2.4 Research Gap

Despite these design considerations, it is still unclear how TAR learning environments compare to less complex smartphone-only and AR-only learning environments with regard to intuitive use and user experience. The higher complexity of interactions could reduce the learning effectiveness of TAR learning environments in comparison to AR-only and smartphone-only approaches. Also, it could reduce the feasibility of integrating TAR learning environments in lessons. Educators do only have a limited amount of time per lesson and hence there often is not much room for an introduction of a complicated learning environment.

3 HORST – THE TEACHING FROG

To compare the three technologies, i.e., smartphone, AR, and TAR, we selected the gamified TAR learning environment *Horst – The Teaching Frog* (*Horst-TAR*) [39] and developed an alternative smartphone-only (*Horst-S*), and AR-only (*Horst-AR*) version (see Figure 1). *Horst-TAR* targets the learning of a frog's anatomy by simulating the dissection of a frog (see Figure 2). Similar to other digital simulations, e.g., *Digital Frog* [14], *Froggipedia* [34], and *Frog Dissection* [11], *Horst-TAR* is an alternative to the questionable approach of dissecting real animals [5, 16]. Research demonstrates a more effective learning using digital simulations instead of real dissections [50].

Using *Horst-TAR*, learners can virtually dissect a frog, closely inspect each organ, learn about their functions, and test their content knowledge in a quiz. The learning environment is a supplementary material for sixth grade biology lessons taught at secondary schools that deal with the amphibian anatomy [18]. The encoded learning contents, e.g., the information about the organs and the procedure of a frog dissection, are based on a typical biology schoolbook [17] and an online frog dissection guide [46].

The design of *Horst-TAR* followed the design considerations in section 2. In particular, interactions with the smartphone application and the tangible objects are designed to be single-handed. This fulfills the core requirement of an overall two-handed design.

3.1 Virtual Frog Dissection

Horst-TAR uses a large, but visually realistic soft toy of a frog as the basis. The belly of the plush frog is a pouch featuring a zipper. By dissecting the frog, i.e., opening the zipper, tangible markers become visible. The paper-card-based markers are attached to the frog using a small piece of velcro, thus making them extractable. Each marker represents an individual organ. Detecting them with the smartphone's camera places a 3D model of each organ above the respective

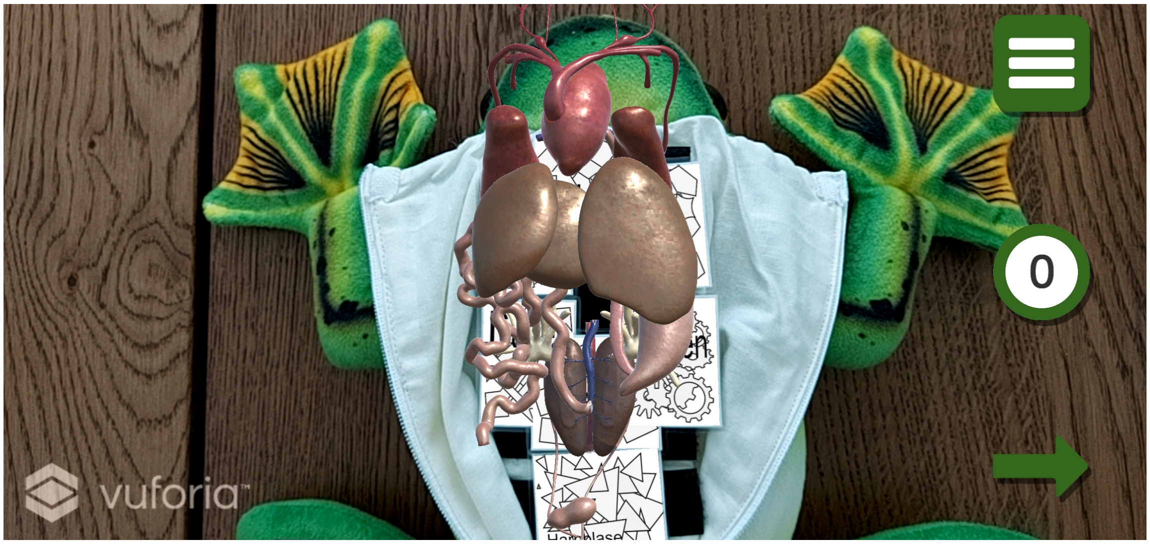

Fig. 2. When detecting the markers with a smartphone, *Horst-TAR* displays 3D models of the organs.

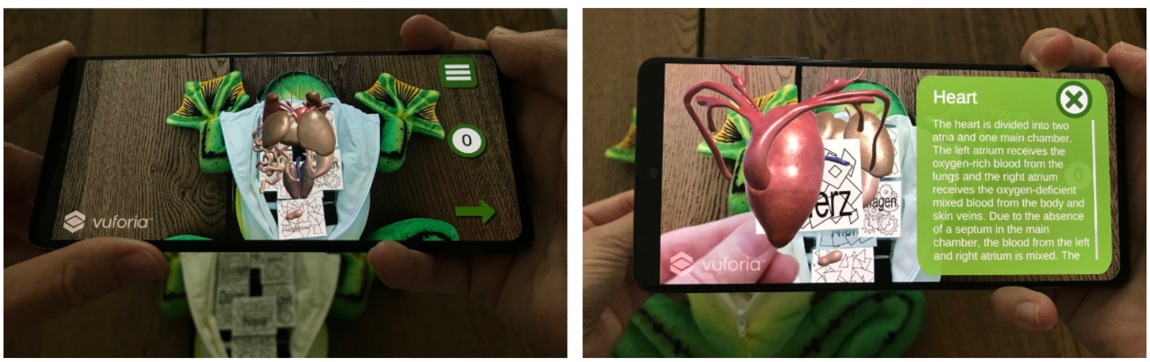

Fig. 3. *Horst-TAR* enables learners to extract individual organ markers and to view them from different angles in a direct way. When tapping a 3D model, the application shows biological information about the respective organ.

target as Figure 2 displays. This enables learners to visualize the internal structure of the animal. Extracting a marker allows for an up close inspection of the organ. Learners can tap on a 3D model of an organ to receive further biological details about it (see Figure 3).

Horst-TAR provides two dissection modes. The *free mode* allows learners to freely examine the frog and its organs without providing any guidance. The *assisted mode* guides a learner through the procedure of a dissection. This mode further provides additional information concerning the frog's anatomy. Guided by the learning environment, learners must find, extract, and open the biological information of an organ to proceed to the next organ. The TAR 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. The application prohibits to skip steps.

Fig. 4. The learning environment provides various achievements to create an incentive for a repetitive use. Achievements are well-defined challenges.

The TAR learning environment further includes a quiz to support the learning process. Learners can assess their learning progress and identify potential knowledge gaps. The quiz informs learners about the correctness of their inputs in an audiovisual way. *Horst-TAR* rewards a correct answer with a quack sound and highlights the selected answer in green. In the case of a wrong answer, the learning environment plays the sound of a buzzer and highlights the learner's selection in red. While *Horst-TAR* removes a correctly answered question from the list of upcoming questions, it returns a wrongly answered question to the list. A learner finishes the quiz when all questions are answered correctly. In its current form, the quiz includes 16 questions testing a user's content knowledge about a frog's anatomy. In addition, *Horst-TAR* features a gamification system, i.e., highscores and an achievement system. Learners earn highscore-points by either finishing an assisted dissection or by performing well in the quiz. In contrast, the achievement system presents well-defined challenges to the users as depicted in Figure 4.

3.2 Versions for Smartphone and Handheld AR

To evaluate the effects of TAR, we further developed two additional learning environments that differ in the technology used but provide the same learning content. *Horst-AR* implements the same features as *Horst-TAR* but is based on a 2D-image printout instead of the plush frog as displayed in Figure 1 and Figure 5. The 2D-image printout features seven fixed markers for the individual organs. By touching a 3D model on the smartphone, the application extracts the respective organ and displays its biological information. A learner can further inspect and rotate the organ by dragging the finger around. A subsequent tab on the dissection-board places the organ on it. This removes the organ from the frog completely. In this way, *Horst-AR* replaces the physical aspect of extracting the organs of *Horst-TAR* by a button-press.

Also, we developed a smartphone-only learning environment, i.e., *Horst-S*. *Horst-S* features no AR functions and directly displays 3D models of the frog and its organs as shown in Figure 6. Implementing the same interactions as *Horst-AR*, learners can dissect the virtual frog using touch and drag gestures. Thus, *Horst-S* represents the baseline and

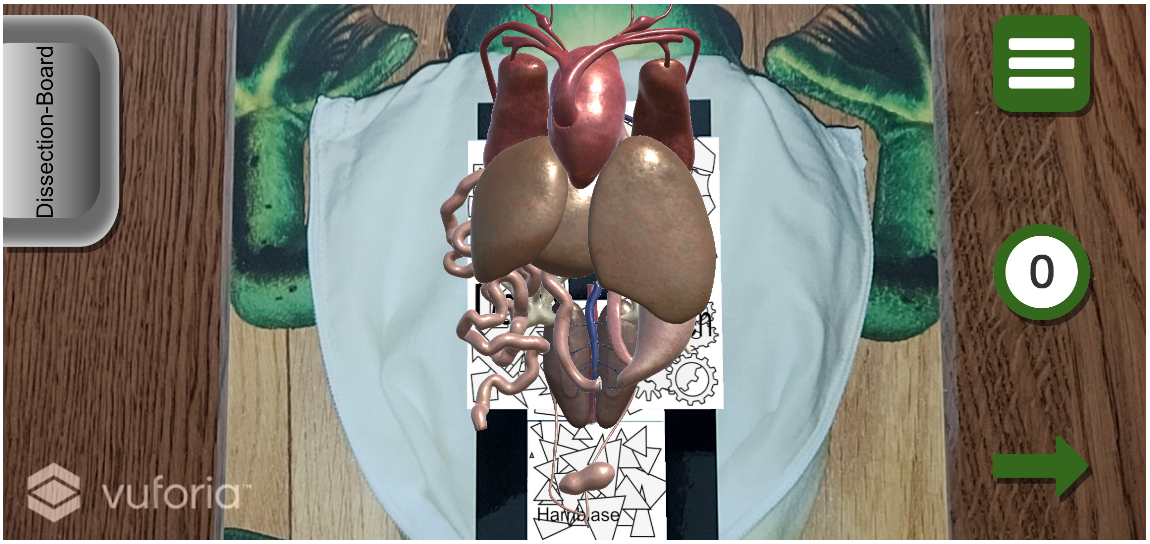

Fig. 5. *Horst-AR* implements a 2D-printout of the frog and follows an AR-only approach. Learners dissect the frog using interactions with the smartphone application.

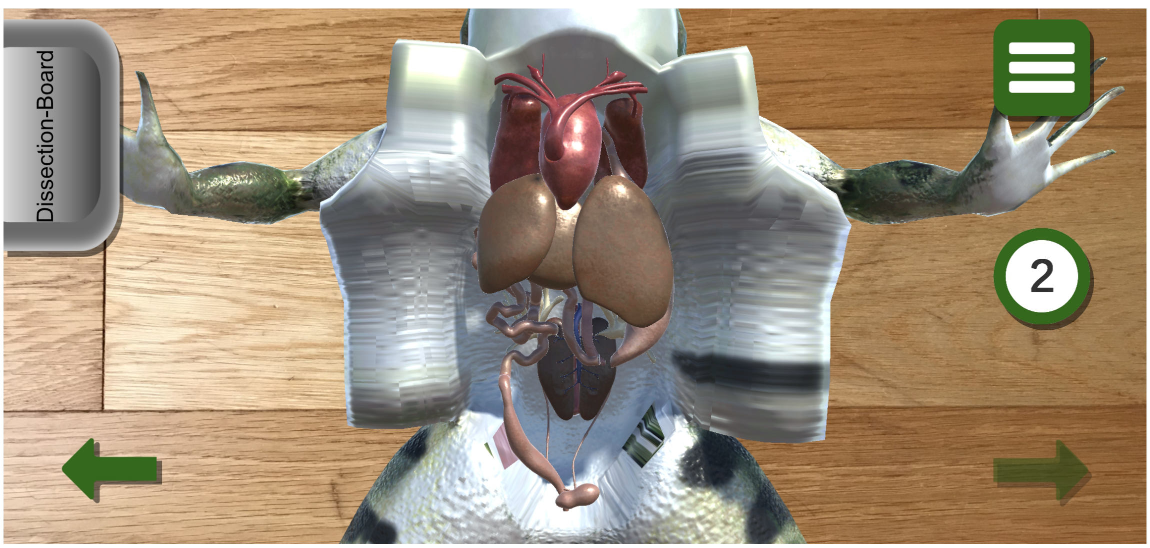

Fig. 6. *Horst-S* allows for a frog dissection on smartphones, thus presenting the de-facto standard for mobile learning.

allows for an investigation of the effects of AR and TAR. The differences between the three versions only apply to the interactions of the dissection. We kept all other features, e.g., the quiz and the achievement system, the same.

3.3 Technology

The application was developed with *Unity* in the version 2019.3.9f1 [48] and targets smartphones with *Android* from version 4.4 and higher. We implemented the AR functions using the *Vuforia Engine* in the version 8.5.9 [43]. The 3D models of the organs and the frog in *Horst-S* are from the *Frog anatomy* package [8].

4 STUDY DESIGN

The overall goal of our research is to investigate the effects of using TAR for educational purposes. This requires a twofold approach. First, TAR learning environments should be validated with respect to their usability in comparison to de-facto state-of-the-art approaches. Second, the systems should be compared concerning the learning effectiveness. This second step requires a comparable usability of all systems to reduce potential confounds caused by the interaction with the system. For instance, a difference in the efficiency of a modality's interaction techniques can affect the learning effectiveness, e.g., fewer exercises can be solved within the same amount of time [41]. By fulfilling this research agenda, insights into the general feasibility of TAR learning environments can be inferred. The present study focuses on the comparison of the three *Horst* versions to (1) gain insights into the intuitive use and user experience of TAR learning applications and to (2) test the usability of the systems to prepare the future learning effectiveness study.

We assume the following hypotheses based on the analysis of the theoretical work in section 2 and our design considerations in section 3.

- H1 When designing for two-handed interactions, TAR yields a similar intuitive use to AR-only and smartphone-only.
- H2 The provision of tangible objects results in a higher user experience in comparison to AR-only and smartphone-only applications.

To answer these hypotheses and to compare the three versions, we conducted a user study utilizing a within subjects design. The participants used the three versions in counterbalanced order. For each of the systems, participants completed an assisted dissection. We also included the quiz in the usability evaluation to sanitize it for the future learning effectiveness study. It supports the learning process and hence is an integral part of the learning environment. As we were only interested in its perceived intuitive use and user experience, we reduced the number of questions to six to limit the time participants need to complete it. As the quiz is the same in all conditions, participants only completed it once after finishing their first dissection using their initial version of *Horst*.

4.1 Measures

We used the following measures to compare the three versions.

4.1.1 Intuitive Use. With the *Questionnaire for the subjective consequences of intuitive use* (*QUESI*) [37], we measured the perceived intuitive use. The questionnaire consists of the subscales of subjective mental workload, perceived achievement of goals, perceived effort of learning, familiarity, and perceived error rate. The scales range from 1 to 5 (5 = high intuitiveness). Participants filled in the questionnaire at the end of each dissection and the quiz.

4.1.2 User Experience. We measured the perceived user experience using the *User Experience Questionnaire* (*UEQ*) [31]. The questionnaire consists of six subscales, i.e., attractiveness (six items), perspicuity, dependability, efficiency, novelty, and stimulation (four items each). Each item consists of two opposite adjectives (e.g., annoying and enjoyable or boring and exciting) and a 7-point Likert scale, thus allowing participants to express their tendency towards the two extremes. Participants filled in the questionnaire at the end of each dissection and the quiz.

4.1.3 Preference. At the end of the experiment, we asked the participants to rank the three versions according to their preference. Each version could only be assigned to one of the three possible ranks. We also encouraged the participants to reason their preference.

4.1.4 Quiz Result. Although we only focused on the usability of the quiz, we logged the total number of a participant's wrong answers in the quiz. The analysis of these results can give a preliminary insight into the learning effectiveness of *Horst*.

4.2 Apparatus

The experimental setup consisted of a desk, an office chair, a cabinet, and a computer for filling in the questionnaires. For playing the three versions, we used a Samsung S10-Lite smartphone. Instead of orally instructing the participants, we used the questionnaire system to display experimental instructions and hence to guide the participants through the study. It instructed them about the next application that needed to be started on the smartphone and the location of the relevant asset, i.e., the plush frog or the 2D-image printout. We made this choice to ensure that all participants received the exact same information. We stored the plush frog and the printout of the frog in separate drawers of the cabinet. This prohibited the participants from seeing these assets before they actually needed them for the respective version of the learning application.

4.3 Procedure

The study took place during the COVID-19 pandemic. To ensure for protection and hygiene, we took the following precautions. (1) Each participant had to disinfect their hands before and after the study, constantly wear a mask as well as latex gloves when interacting with the plush frog, and report whether they stayed in a risk area or show signs of an illness. (2) The experimenter had to disinfect their hands, constantly wear a mask, and daily report whether they show signs of an illness. (3) The experimenter and the participant had to keep at least a distance of 1.5 meters. (4) All touched surfaces and used devices, e.g., smartphone and keyboard, had to be cleaned with a disinfectant product after each experimental trial. (5) The laboratory had to be ventilated for at least 15 minutes after each experimental trial.

After welcoming the participants, we explained them the goal of the experiment as well as the protection and hygiene measurements. The participants then gave their written consent to voluntarily participate in the study and took a pair of fresh latex gloves. We decided to use the latex gloves as cleaning the plush frog with a disinfectant product would have ruined it. Afterwards, the participants took seat at the desk and started to fill in the demography questionnaire as well as to follow the displayed experimental instructions. After completing an assisted dissection or the quiz, the participants filled in a mid-questionnaire consisting of the QUESI and the UEQ to rate the just used version. The participants completed the quiz only once after finishing the first dissection and mid-questionnaire. Once the participants completed the last mid-questionnaire, they ranked the three versions of the learning application according to their preference. Finally, we thanked them for participating in the experiment.

4.4 Participants

We recruited our participants from the students enrolled at the University of Würzburg. As a reward, they received credits mandatory for obtaining their program of study's degrees. In total, 30 participants (25 females, 5 males) took part in the experiment. The mean age of the participants was 22.37 (SD = 3.50) years. All participants were native speakers and none had an impaired vision that was not corrected during the experiment. With regard to frequency of computer usage, 19 participants reported a frequent use per day, 9 a use on a daily basis, 1 a use on a weekly basis, and 1 a use on a monthly basis. With regard to playing computer games, 16 participants reported to never play video games, 8 to play on a monthly basis, 4 on a weekly basis, and 2 on a daily basis. 21 participants reported to use mobile applications multiple times per day, 4 on a daily basis, 4 on a weekly basis, and 1 never used mobile applications before. With regard to AR mobile applications, 19 participants reported to never use AR, 8 on a monthly basis, and 3 on a weekly basis.

5 RESULTS

We tested for sphericity by computing Mauchly's W. In the case of a violation, we applied a Greenhouse-Geisser correction. To compare the results, we computed repeated measures analysis of variance (RM-ANOVA). Follow-up tests were pairwise comparisons with Bonferroni-Holm-adjustments. Effect sizes were determined by computing η^2. We compared the results of the quiz by computing an analysis of variance (ANOVA) after testing for homogeneity of variance using a Levene's test. The descriptive statistics are displayed in Table 1.

Table 1. Descriptive statistics; $N = 30$. Values are $M(SD)$.

Scale	Horst-S	Horst-AR	Horst-TAR	Quiz
QUESI total score	4.486 (0.460)	4.362 (0.631)	4.486 (0.366)	4.488 (0.477)
Subjective mental workload	4.389 (0.561)	4.222 (0.708)	4.344 (0.483)	4.556 (0.596)
Perceived achievement of goals	4.667 (0.420)	4.556 (0.713)	4.667 (0.429)	4.244 (0.705)
Perceived effort of learning	4.456 (0.590)	4.367 (0.697)	4.522 (0.452)	4.656 (0.514)
Familiarity	4.389 (0.613)	4.233 (0.759)	4.333 (0.479)	4.456 (0.543)
Perceived error rate	4.550 (0.607)	4.467 (0.765)	4.600 (0.593)	4.550 (0.578)
UEQ				
Attractiveness	1.700 (0.976)	1.678 (1.172)	2.067 (0.909)	1.539 (1.031)
Perspicuity	2.242 (0.696)	1.950 (0.794)	2.158 (0.638)	2.358 (0.649)
Dependability	1.683 (0.707)	1.475 (0.826)	1.500 (0.692)	1.692 (0.712)
Efficiency	1.867 (0.779)	1.500 (0.991)	1.758 (0.795)	1.967 (0.571)
Novelty	1.025 (1.335)	1.642 (1.035)	2.042 (0.994)	0.500 (1.369)
Stimulation	1.542 (0.825)	1.808 (0.858)	2.242 (0.756)	1.467 (0.995)
Quiz number of wrong answers	5.250 (3.882)	6.700 (3.561)	5.800 (2.530)	—

5.1 Intuitive Use

We found no significant difference between the three learning environments for the subscales of *subjective mental workload*, $W = .943$, $p = .441$; $F(2, 58) = 0.703$, $p = .499$; $\eta^2 = .024$, *perceived effort of learning*, $W = .813$, $p = .055$; $F(2, 58) = 0.498$, $p = .610$; $\eta^2 = .017$, and *perceived error rate*, $W = .942$, $p = .435$; $F(2, 58) = 0.313$, $p = .733$; $\eta^2 = .011$. Sphericity was violated for *perceived achievement of goals*, $W = .733$, $p = .013$, and *familiarity*, $W = .777$, $p = .029$, hence we applied a Greenhouse-Geisser correction. Comparing the results, we found no significant difference for *perceived achievement of goals*, $F(1.578, 45.760) = 0.557$, $p = .537$; $\eta^2 = .019$, and *familiarity*, $F(1.636, 47.430) = 0.462$, $p = .594$; $\eta^2 = .016$.

Since only 25 participants reported an at least daily use of smartphones, the remaining participants might have skewed the results due to an effect of novelty. Thus, we additionally compared the three conditions with regard to the *QUESI total score* for the 25 experienced participants to test for such an effect of novelty. We found no significant difference between the conditions, $F(2, 48) = 0.712$, $p = .496$.

Finally, the quiz received high ratings of perceived intuitive use with all subscales and the total score above 4.

5.2 User Experience

Table 2. User experience follow-up pairwise Bonferroni-Holm-adjusted comparisons, significant differences are highlighted; $N = 30$.

Subscale	**Condition 1**	**Condition 2**	***t***	***p* (adj.)**
Attractiveness	Horst-S	Horst-AR	0.148	0.883
	Horst-S	**Horst-TAR**	**-2.880**	**0.022**
	Horst-AR	**Horst-TAR**	**-2.580**	**0.023**
Efficiency	Horst-S	Horst-AR	2.500	0.055
	Horst-S	Horst-TAR	0.859	0.397
	Horst-AR	Horst-TAR	−1.580	0.188
Novelty	**Horst-S**	**Horst-AR**	**-2.510**	**0.018**
	Horst-S	**Horst-TAR**	**-4.150**	**< 0.001**
	Horst-AR	**Horst-TAR**	**-2.860**	**0.012**
Stimulation	Horst-S	Horst-AR	−1.400	0.172
	Horst-S	**Horst-TAR**	**-4.250**	**< 0.001**
	Horst-AR	**Horst-TAR**	**-2.730**	**0.016**

We found a significant difference for *attractiveness* between the three version, $W = .956$, $p = .536$; $F(2, 58) = 4.666$, $p = .013$; $\eta^2 = .139$. Follow-up tests revealed a significantly higher rating for *Horst-TAR* than *Horst-S* and for *Horst-TAR* than *Horst-AR*. Table 2 provides an overview of the follow-up comparisons. For *perspicuity*, we found no significant difference between the three version, $W = .984$, $p = .795$; $F(2, 58) = 1.797$, $p = .175$; $\eta^2 = .058$. We found no significant difference for *dependability* between the three version, $W = .862$, $p = .125$; $F(2, 58) = 1.142$, $p = .326$; $\eta^2 = .038$. For *efficiency*, we found a significant difference between the three version, $W = .914$, $p = .283$; $F(2, 58) = 3.316$, $p = .043$; $\eta^2 = .103$. Follow-up tests, in contrast, revealed no significance between the conditions. Sphericity was violated for *novelty*, $W = .664$, $p = .003$, hence we applied a Greenhouse-Geisser correction. We found a significant difference for this subscale between the three version, $F(1.497, 43.423) = 11.244$, $p < .001$; $\eta^2 = .279$. Follow-up tests revealed a significantly higher rating for *Horst-TAR* than *Horst-S*, for *Horst-TAR* than *Horst-AR*, and for *Horst-AR* than for *Horst-S*. We found a significant difference for *stimulation* between the three version, $W = .946$, $p = .461$; $F(2, 58) = 8.444$, $p < .001$; $\eta^2 = .226$. Follow-up tests revealed a significantly higher rating for *Horst-TAR* than *Horst-S* and for *Horst-TAR* than *Horst-AR*.

Similar to the QUESI results, we also checked for a potential effect of novelty by additionally comparing the three conditions with regard to the significant subscales for the 25 experienced participants only. We still found significant differences for *attractiveness* ($F(2, 48) = 3.809$, $p = .029$), *novelty* ($F(2, 48) = 9.676$, $p < .001$), and *stimulation* ($F(2, 48) = 7.096$, $p = .002$).

Finally, the quiz received high ratings for *perspicuity*, $M = 2.358$, $SD = 0.649$, and *efficiency*, $M = 1.967$, $SD = 0.571$. Except for *novelty*, the other subscales were in the middle of the positive range of the scale. Participants gave only a slightly positive rating on the *novelty*.

5.3 Preference

The participants preferred *Horst-TAR* the most, followed by *Horst-AR*, and *Horst-S*. For rank 1, 18 (60%) participants selected *Horst-TAR*, 6 (20%) selected *Horst-S*, and 6 (20%) selected *Horst-AR*. For rank 2, 14 (47.67%) participants selected *Horst-AR*, 8 (27.67%) selected *Horst-S*, and 8 (27.67%) selected *Horst-TAR*. For rank 3, 16 (53.33%) participants selected *Horst-S*, 10 (33.33%) selected *Horst-AR*, and 4 (13.33%) selected *Horst-TAR*.

Six participants praised *Horst-TAR* for using tangible objects and hence the resulting direct combination of virtuality with reality. Six participants further enjoyed the haptic aspect and called *Horst-TAR* innovative. Also, five participants named the perceived realism of dissecting the frog by extracting its organs as a positive aspect. Four participants reported that the plush frog elicits supportive aspects for learning. As negative comments, four participants reported issues with either the scanning or the procedure of extracting the organ markers. Two additional participants were concerned about the realism of the plush frog. Concerning positive comments about *Horst-AR*, three participants found the experience of AR interesting and further three participants praised the direct inspection of the virtual organs. As negative aspects, two participants questioned the benefits of using AR for learning about the anatomy of a frog. Also, one participant disliked the flat 2D-image printout and one participant found the interaction with the AR learning environment complicated. Five participants highlighted the lack of additional components of *Horst-S* as a positive aspect. It can be used everywhere and at every time as long as a smartphone is provided. Four participants further praised the realistic 3D model of the frog as a positive aspect. Concerning negative aspects, five participants called *Horst-S* a common and hence less innovative approach.

5.4 Quiz Result

The logging of the quiz results failed for two participants of the *Horst-S* condition. Computing a Levene's test revealed no violation of the homogeneity of variance, $p = .780$. The comparison of the three learning environments revealed no significant differences concerning the results in the quiz, $F(2, 25) = .440$, $p = .649$.

6 DISCUSSION

The present study compares the differences in perceived intuitive use and user experience evoked by different technologies of knowledge presentation in an educational context. In particular, we compared these qualities between a smartphone-only, AR-only, and TAR version of a biology learning environment. The versions only differed in the technology used, but presented the same learning content.

6.1 Intuitive Use

With regard to intuitive use, the three versions of *Horst* did not differ significantly. This outcome is an important finding for educators and developers. Tangible objects can provide additional educational aspects, e.g., physical properties, and support a direct interaction with a learning content [3, 4]. Our result indicates that all technologies examined can be used in an educational context without compromising the intuitive use.

Research demonstrates a higher intuitiveness when using TUIs instead of GUIs [7]. However, mobile applications targeting smartphones should be designed to only require simple interactions [23]. A user should be able to execute interactions single-handedly to maximize usability [32]. This becomes even more relevant when using AR and TAR. Here, especially when using TAR, users may also have to simultaneously manipulate tangible objects. This could lead to a reduced intuitiveness of TAR learning environments. Hence, TAR applications should be designed for two-handed interactions. When following this design goal, our results indicate that using mobile TAR is not decremental for intuitiveness. **Thus, our results confirm H1.**

6.2 User Experience

Concerning user experience, our results suggest AR and TAR to be beneficial compared to the de-facto standard of using smartphone-only applications. This supports previous research showing an improved usability of TAR applications [1]. Both AR versions of *Horst* were superior to the smartphone-only version with respect to the subscale of *novelty*. Nineteen participants reported to have never used AR before. As a result, experiencing this technology for the first time throughout the study might have contributed to the significant differences. However, this might also be true for pupils who encounter AR for the first time in an educational context, e.g., when learning about the anatomy of vertebrates. Thus, the novelty of using AR learning environments could be beneficial for the overall motivation and perceived quality of the lessons. However, the effect of novelty can decline over time. This would then reduce or cease the novelty benefits of using AR or TAR in teaching concepts. Hence, this result might only be true for our current sample.

TAR also yielded positive effects on other subscales of the UEQ. In particular, our results indicate a significant positive effect of TAR with regard to the subscales of *attractiveness* and *stimulation* in comparison to smartphone-only and AR-only. This suggests that TAR generally improves the user experience of learning environments. The positive effects of TAR could be influenced by the perceived novelty of this approach. However, the lack of a significant difference between *Horst-AR* and *Horst-S* with regard to these two qualities contradicts this assumption. In addition, we tested for an effect of novelty by focusing on participants with at least a daily use of mobile apps during our statistical analysis. This accounted for novelty effects evoked by the use of a mobile device during our experiment. In this follow up analysis, we still found significant differences concerning *attractiveness* and *stimulation*. **Thus, our results confirm H2.**

The lack of a statistically significant difference between the three tested learning environments with regard to *efficiency* provides further insights into our design of the TAR interaction techniques. *Horst-TAR* requires the simultaneous interaction with a smartphone and the tangible elements. In contrast, *Horst-S* merely requires the interaction with a smartphone. Despite this higher complexity, the participants did not experience a reduced efficiency of *Horst-TAR*. This result supports our theoretical considerations and design choices of implementing two-handed interactions.

Hence, our results suggest that TAR greatly increases the user experience. When integrating TAR learning environments in lessons, the positive experience could lead to a higher learning motivation and overall increased enjoyment [21]. The second step of our research agenda targets the evaluation of the learning motivation and learning effectiveness when using either one of the three versions. We plan to measure the learning effectiveness using a pre-test post-test study design. Before the first exposure to either one of the three learning environments or a traditional schoolbook approach, we intend to assess the participants' knowledge about the anatomy of a frog by a multiple-choice test. Subsequently, the participants will use a randomly assigned method to acquire new knowledge about a frog's anatomy. Finally, we will re-test the participants' knowledge using a different multiple-choice test of the same difficulty to the first one. In addition, we will assess the learning motivation of the participants. We hypothesize for this experiment that all learning environments will yield a similar learning effectiveness but elicit a higher learning motivation compared to the traditional approach. Depending on the current situation concerning COVID-19 at the time of the study, we plan to conduct the study with sixth grade students at local schools.

Since attractiveness and stimulation might also decline over time or be influenced by an effect of novelty, it is also important to analyze the longterm effects of using TAR learning environments. To gain preliminary insights into these longterm effects, we intend to continue the training period after the post-test knowledge assessment. We will ask the participants to repetitively use their respective learning environment and assess the user experience after each exposure.

6.3 Preference

The positive effect of TAR on the user experience is backed by the outcome of the preference rating. Participants preferred TAR over AR and smartphone as well as AR over smartphone. The participants' comments about the three learning environments support the user experience results even further. Providing TAR enabled a haptic and realistic perception of the virtual dissection. Participants also praised this technology for the direct combination of virtuality and reality. This supports previous research showing the benefits of achieving a physical aspect [7] and hence a direct and haptic interaction [4] with a learning environment. While the participants' reasoning confirmed the advantages of a direct interaction with the learning content [3], they also questioned the use of AR for educational purposes. This negative aspect could be explained by the perceived advantages of *Horst-S*. The smartphone-only learning environment yielded a realistic visualization of the frog without requiring any additional materials. Taken together with the findings described above, we conclude that TAR learning environments provide despite their higher complexity a similar intuitiveness while outperforming other technologies with regard to user experience. This suggest that TAR learning environments could greatly improve the learning motivation.

6.4 Quiz Result

The analysis of the quiz revealed no significant difference between the three learning environments. It suggests that the learning environments elicit a similar learning effectiveness when completing the assisted dissection for the first time. This result is promising as it indicates that none of the tested technologies causes a negative effect on the learning effectiveness. Also, it is an important insight for educators as they can safely integrate all tested modalities in teaching concepts. However, we explained the participants the goal of our study before they started the first experimental trial. As a result of this, they might have focussed more on the actual interaction with their initial version of *Horst* than on the presented learning content. The participants might have quickly skipped through the remaining organs after having developed an understanding for the respective interaction techniques. Also, to limit the time to complete the quiz, we reduced the number of questions, thus no longer assessing all aspects of the learning content. Therefore, it is important to conduct a learning effectiveness study as future work to compare the three modalities with regard to this quality.

6.5 Implications

Our results might be of high interest for educators and developers. Especially in the case of extended phases of remote teaching, it can be vital to provide learning environments that are intuitive to use, elicit a high user experience, evoke a high motivation, and present the learning content in a direct way. Based on the theoretical background and our results, using TAR learning environments has the potential to fulfill these requirements. Educators can safely integrate TAR learning environments in their teaching concepts to achieve a direct interaction with the learning contents. This especially can be beneficial when the learning contents are difficult to demonstrate otherwise. However, especially in a situation of remote teaching, this can also be complicated as learners need additional tangible objects besides their smartphones. In such a scenario, using smartphone-only or AR-only learning environments is still a valid alternative. While eliciting a lower attractiveness and stimulation, they lack the requirement of physical objects. Table 3 summarizes our considerations.

Table 3. Comparison of smartphone, AR, and TAR learning environments.

	Smartphone-only	**AR-only**	**TAR**
Accessibiliy	Highest	Moderate	Lowest
Material	Smartphone	Smartphone & marker printouts	Smartphone & tangible markers
Design	Mobile-specific design	Mobile- & AR-specific design	Mobile-, AR- & single-handed tangible design
	Touch manipulation on smartphone	Touch manipulation on smartphone	Touch manipulation on smartphone & haptic manipulation by tangibles
	Single-handed operability	Single-handed operability	Two-handed operability
Integration	Remote & classroom teaching	Remote & classroom teaching	Classroom teaching

To provide these TAR learning environments, we suggest that they should be designed for two-handed interactions, i.e., single-handed interactions for controlling the smartphone and the tangible objects, to maximize intuitive use and efficiency. In addition, to maximize user experience, interactions with the learning contents should be implemented in a direct, rich, and physical way [22]. This should lead to an immediate combination of the virtual information with the real-world and hence in a potentially higher acceptance as well as learning motivation. Also, by directly interacting with haptic objects, learners can perceive a higher degree of realism. By following these two guidelines, TAR learning environments should allow learners to easily apply and observe changes to the learning contents.

6.6 Limitations

We conducted our study during the COVID-19 pandemic. This led to measurements to protect the hygiene and safety. Participants had to wear latex gloves during the experiment. Wearing these gloves could have negatively influenced the perception of the plush frog and the extractable organ markers. In addition, it could have resulted in an unusual feel when interacting with the smartphone in the TAR condition. As a result, the user experience might have been negatively influenced by this measurement. However, despite this measurement, *Horst-TAR* outperformed the other two conditions.

A second limitation is our sample. The learning environment targets students of the sixth grade. Unfortunately, conducting studies at local schools was not possible due to the COVID-19 pandemic. Hence, our results can only indicate a general higher user experience when using TAR but not provide insights into the perception of TAR of sixth grade students. In addition, the majority of the participants experienced AR for the first time. This might have skewed the results of the user experience evaluation.

Our sample further presents a limitation concerning the results of the quiz. All participants were students at university level and hence already learned about the anatomy of vertebrates while attending high-school. In this way, using *Horst* most likely helped them to merely recall this knowledge instead of actually acquiring new knowledge about the anatomy of a frog. Thus, the quiz might have not primarily gauged the learning effectiveness but measured the potential to recall already acquired knowledge.

Finally, the plush frog is a very playful and abstract representation of a frog. Utilizing a more realistic model of a frog with respect to size and surface texture could greatly improve the educational effect.

7 CONCLUSION

In the present study, we investigated the effects of varied mobile knowledge presentation techniques in an educational context. In particular, we compared the perceived intuitive use and user experience of using three versions of a learning environment. The versions only differ with respect to the technology used, i.e., smartphone-only, AR-only, and TAR.

Our results show that all technologies examined can yield a similar perceived intuitive use. Based on this outcome, the three technologies can safely be integrated in learning environments without compromising the intuitiveness. For TAR, we recommend to design for two-handed interactions to enhance intuitive use and efficiency. The smartphone application and the tangible objects should be single-handedly usable to allow for an easy interaction. With regard to user experience, TAR indicated to be the most attractive and stimulating in contrast to AR-only and smartphone-only. This suggests a high benefit of using TAR for learning environments.

Future work should focus on analyzing the effects of using TAR learning environments compared to AR-only and smartphone-only approaches with regard to learning effectiveness, motivation, and cognitive load. A further research direction should be to improve the tangible objects of *Horst-TAR* by creating image targets in form of the actual organs and exchanging the plush frog with a more realistic model. Finally, research should analyze potential changes in the perceived user experience in a longterm study.

ACKNOWLEDGMENTS

We like to thank David Schraudt for his assistance in developing the three different learning environments.

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

REFERENCES

[1] Alissa Antle and Alyssa Wise. 2013. Getting down to details: Using learning theory to inform tangibles research and design for children. *Interacting with Computers* 25 (01 2013), 1–20. https://doi.org/10.1093/iwc/iws007
[2] Ronald T Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (1997), 355–385.
[3] Jorge Bacca, Silvia Baldiris, Samon Fabregat, Sabine Graf, and Kinishuk. 2014. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educational Technology & Society* 17, 4 (2014), 133–149.
[4] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. 2018. 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 (2018), 457–467.
[5] Jonathan Balcombe. 2013. Alternatives to dissection. In *The Global Guide to Animal Protection*, Andrew Linzey (Ed.). University of Illinois Press, 271–272.
[6] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. 1993. Toolglass and Magic Lenses: The See-through Interface. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (Anaheim, CA) *(SIGGRAPH '93)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/166117.166126
[7] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2008. Tangible augmented reality. *ACM SIGGRAPH ASIA 2008 Courses* (01 2008). https://doi.org/10.1145/1508044.1508051
[8] B'in Games. 2018. Frog anatomy. https://assetstore.unity.com/packages/3d/characters/animals/frog-anatomy-113389.
[9] T Chandrasekera and S Yoon. 2015. 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*. 6–8. https://doi.org/10.1109/ISMAR-MASHD.2015.18
[10] Dennis Charsky. 2010. From Edutainment to Serious Games: A Change in the Use of Game Characteristics. *Games and Culture* 5, 2 (2010), 177–198. https://doi.org/10.1177/1555412009354727
[11] GP Strategies Corporation. 2020. Frog Dissection. available at https://apps.apple.com/us/app/frog-dissection/id377626675.
[12] Sara de Freitas and Fotis Liarokapis. 2011. Serious Games: A New Paradigm for Education? In *Serious Games and Edutainment Applications*, Minhua Ma, Andreas Oikonomou, and Lakhmi C Jain (Eds.). Springer-Verlag, London, 9–23. https://doi.org/10.1007/978-1-4471-2161-9{_}2
[13] P Diegmann, M Schmidt-Kraepelin, S Van den Eynden, and D Basten. 2015. Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. In *Proceedings of the 12th International Conference on Wirtschaftsinformatik*. Osnabrück, Germany, 1542–1556.
[14] Inc. Digital Frog Internation. 2020. Digital Frog. available at https://www.digitalfrog.com.

[15] Tristan C Endslay, Kelly A Sprehn, Ryan M Brill, Kimberly J Ryan, Emily C Vincent, and James M Martin. 2017. Augmented Reality Design Heuristics: Designing for Dynamic Interactions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 2100–2104. https://doi.org/10.1177/1541931213602007

[16] Kenneth R Fleischmann. 2003. Frog and Cyberfrog are Friends: Dissection Simulation and Animal Advocacy. *Society & Animals* 11, 2 (2003), 123–143.

[17] Thorsten Fraterman, Ramón Gomez-Islinger, Dietmar Kalusche, Alexander Röhrer, and Marianne Walcher. 2017. *PRISMA Biologie 6* (1st ed.). Ernst Klett Verlag, Stuttgart/Leipzig.

[18] Staatsinstitut für Schulqualität und Bildungsforschung München. 2020. LehrplanPLUS. available at https://www.lehrplanplus.bayern.de/.

[19] M Gervautz and Dieter Schmalstieg. 2012. Anywhere Interfaces Using Handheld Augmented Reality. *Computer* 45, 7 (2012), 26–31. https://doi.org/10.1109/MC.2012.72

[20] Joanne Gikas and Michael M Grant. 2013. Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education* 19 (2013), 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002

[21] Julián Esteban Gutiérrez Posada, Elaine C S Hayashi, and M. Cecília C Baranauskas. 2014. On Feelings of Comfort, Motivation and Joy that GUI and TUI Evoke. In *Design, User Experience, and Usability. User Experience Design Practice*, Aaron Marcus (Ed.). Springer International Publishing, Cham, 273–284.

[22] Steve Hinske, Marc Langheinrich, and Matthias Lampe. 2008. Towards Guidelines for Designing Augmented Toy Environments. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems*. 78–87. https://doi.org/10.1145/1394445.1394454

[23] Zhao Huang. 2019. Developing Usability Heuristics for Recommendation Systems Within the Mobile Context. In *Design, User Experience, and Usability. Practice and Case Studies*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, Cham, 143–151.

[24] Epic Games Inc. 2021. Unreal Engine. available at https://www.unrealengine.com/.

[25] Hiroshi Ishii and Brygg Ullmer. 1997. 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)*. Atlanta, USA, 234–241. https://doi.org/10.1145/258549.258715

[26] Mina Khan, Fernando Trujano, Ashris Choudhury, and Pattie Maes. 2018. Mathland: Playful Mathematical Learning in Mixed Reality. In *CHI'18 Extended Abstracts*. Montréal, Canada.

[27] K Kim, Mark Billinghurst, Gerd Bruder, H B Duh, and G F Welch. 2018. 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 (2018), 2947–2962. https://doi.org/10.1109/TVCG.2018.2868591

[28] Andreas Knote, Sabine C Fischer, Sylvain Cussat-Blanc, Florian Niebling, David Bernard, Florian Cogoni, and Sebastian von Mammen. 2019. Immersive Analysis of 3D Multi-cellular In-Vitro and In-Silico Cell Cultures. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. https://doi.org/10.1109/AIVR46125.2019.00021

[29] Andreas Knote, Sebastian von Mammen, YunYun Gao, and Andrea Thorn. 2020. Immersive Analysis of Crystallographic Diffraction Data. In *26th ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3385956.3422097

[30] S M Ko, W S Chang, and Y G Ji. 2013. Usability Principles for Augmented Reality Applications in a Smartphone Environment. *International Journal of Human-Computer Interaction* 29, 8 (2013), 501–515.

[31] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction, Usability Engineering of the Austrian Computer Society on HCI, Usability for Education, and Work (Eds.). Springer-Verlag, 63–76.

[32] Woon-Hyung Lee and Hyun-Kyung Lee. 2016. The usability attributes and evaluation measurements of mobile media AR (augmented reality). *Cogent Arts & Humanities* 3, 1 (2016). https://doi.org/10.1080/23311983.2016.1241171

[33] Lili Liu, Christian Wagner, and Ayoung Suh. 2017. Understanding the Success of Pokémon Go: Impact of Immersion on Players' Continuance Intention. In *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments*, Dylan D Schmorrow and Cali M Fidopiastis (Eds.). Springer International Publishing, Cham, 514–523.

[34] Designmate (I) Pvt. Ltd. 2020. Froggipedia. available at https://apps.apple.com/us/app/froggipedia/id1348306157.

[35] Zhihan Lv, Alex Tek, Franck Da Silva, Charly Empereur-mot, Matthieu Chavent, and Marc Baaden. 2013. Game On, Science - How Video Game Technology May Help Biologists Tackle Visualization Challenges. *PLoS ONE* 8, 3 (2013), e57990. https://doi.org/10.1371/journal.pone.0057990

[36] Jane McGonigal. 2011. *Reality is Broken: Why Games Make Us Better and How They Can Change the World* (1st ed.). Penguin Press, New York.

[37] Anja Naumann and Jörn Hurtienne. 2010. Benchmarks for Intuitive Interaction with Mobile Devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, Lisboa, Portugal, 401–402. https://doi.org/10.1145/1851600.1851685

[38] Sebastian Oberdörfer. 2021. *Better Learning with Gaming: Knowledge Encoding and Knowledge Learning Using Gamification.* Ph.D. Dissertation. University of Würzburg. https://doi.org/10.25972/OPUS-21970

[39] Sebastian Oberdörfer, Anne Elsässer, David Schraudt, Silke Grafe, and Marc Erich Latoschik. 2020. Horst – The Teaching Frog: Learning the Anatomy of a Frog Using Tangible AR. In *Proceedings of the 2020 Mensch und Computer Conference (MuC '20)*. Magdeburg, Germany, 303–307. https://doi.org/10.1145/3404983.3410007

[40] Sebastian Oberdörfer and Marc Erich Latoschik. 2018. 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. https://doi.org/10.1109/VS-Games.2018.8493425

preprint

[41] Sebastian Oberdörfer and Marc Erich Latoschik. 2019. Knowledge Encoding in Game Mechanics: Transfer-Oriented Knowledge Learning in Desktop-3D and VR. *International Journal of Computer Games Technology* 2019 (2019). https://doi.org/10.1155/2019/7626349

[42] Sebastian Oberdörfer and Marc Erich Latoschik. 2019. Predicting Learning Effects of Computer Games Using the Gamified Knowledge Encoding Model. *Entertainment Computing* 32 (2019). https://doi.org/10.1016/j.entcom.2019.100315

[43] PTC, Inc. 2021. Vuforia. available at https://developer.vuforia.com.

[44] Katie Seaborn and Deborah I Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

[45] Sophia C Steinhaeusser, Anna Riedmann, Max Haller, Sebastian Oberdörfer, Kristina Bucher, and Marc Erich Latoschik. 2019. 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.

[46] Home Science Tools. 2020. Frog Dissection Guide Project. available at https://learning-center.homesciencetools.com/article/frog-dissection-project/.

[47] Unity. 2021. Unity. available at https://unity.com.

[48] Unity. 2021. Unity. available at https://unity3d.com/unity/whats-new/2019.3.9.

[49] Penny Ur. 1996. *A Course in Language Teaching*. Cambridge University Press.

[50] Christine Youngblut. 2001. *Use of multimedia technology to provide solutions to existing curriculum problems: Virtual frog dissection*. Ph.D. Dissertation. George Mason University, United States.