Motivational benefits and usability of a handheld Augmented Reality game for anatomy learning

Jonas Blattgerste
Mixed Reality Lab
University of Applied Sciences Emden/Leer
Emden, Germany
jonas.blattgerste@hs-emden-leer.de

Jannik Franssen
Mixed Reality Lab
University of Applied Sciences Emden/Leer
Emden, Germany
jannik.franssen@stud.hs-emden-leer.de

Michaela Arztmann
Department of Education
Utrecht University
Utrecht, Netherlands
m.arztmann@uu.nl

Thies Pfeiffer
Mixed Reality Lab
University of Applied Sciences Emden/Leer
Emden, Germany
thies.pfeiffer@hs-emden-leer.de

***Abstract*—To explore the benefits of scalable, interactive handheld Augmented Reality (AR) serious games for anatomical terminology learning, an AR learning game for German-Latin terminology of bone-structures and areas of the female pelvis was developed and evaluated. The evaluation of the game with 36 midwifery students was conducted with a focus on intrinsic motivation, usability and perceived usefulness of the app. The results indicated strong effect sizes of increased perceived competence and intrinsic motivation, and a medium effect size for increased pressure (Intrinsic Motivation Inventory) after completing the AR game. Furthermore, students perceived the usability of the game as "best imaginable" (contextualization of SUS scores) and strongly agreed that the app is very useful to them.**

***Index Terms*—Augmented Reality, Learning, Serious Game, Medical, Midwifery, Anatomy, Physiology, Pelvis, Bones**

I. Introduction

Approaches for teaching are constantly evolving, and its methods are improving. This is especially true in medical education, where stakes are high and errors can have catastrophic consequences. Flipped classroom methodologies, self-directed learning approaches, case-based teaching, game-based learning, and holistic learning approaches in medical education settings all serve the purpose of engaging students in higher-level thinking and utilizing available time in lectures and practical sessions as effectively and efficiently as possible. Some Mixed Reality technologies, like Augmented Reality (AR), offer a broad set of tools to support these approaches by acting as a simulator, coach or even examiner during learning. They can also motivate through gamification aspects and improve student self-determination. That using Augmented Reality in education can lead to improved learning outcomes and retention, as well as to higher motivational benefits, is already well-known and supported by several systematic reviews and meta-analyses [1]–[3].

Although these results are promising, they are mostly achieved using Head-Mounted Displays (HMDs), projections, or screen-based AR, and a wide-spread curricular adoption still has not occurred. This is partially the case because, to date, those approaches remain challenging to scale because of cost factors as well as the requirement for intensive onboarding and training for both lecturers and students. Current literature also addresses a broad range of topics, and more individual AR learning/training interventions in specifically the medical field are welcomed. Furthermore, the handheld Augmented Reality approaches that exist mostly focus on aspects of 3D visualization and exploration, with only little interactivity [4], [5]. This, in our opinion, limits their benefits when compared to their immersive counterparts, such as HMD-based AR or even Virtual Reality-based approaches. Making use of AR in medical education could be scaled significantly, by extending the applications of Handheld AR beyond 3D visualization and improving upon its potential benefits from engaging interactivity. Here, Bring-Your-Own-Device (BYOD) approaches could be utilized, where students use their own smartphones for AR-based learning instead of institutions buying and maintaining expensive and quickly outdated immersive technology. Arguably, these benefits do not necessarily have to be an increase in retention of learning content when comparing the technology directly to conventional methods such as textbook learning or using index cards. To convince practitioners, it would likely suffice to show that it could improve intrinsic motivation for students to engage with historically "dry" subjects to make its application worthwhile. And compared to maintaining a large collection of 3D plastics, interactive digital 3D models would actually save resources.

The contribution of this paper is two-fold: First, we develop an interactive handheld AR game for anatomical terminology learning in academic midwifery education, where students not only visualize but also actively contextualize German-Latin terminology of the regions and bone structures of the female pelvis. In contrast to previous work, a focus is on achieving this with a scalable, markerless AR implementation on handheld AR devices, while still integrating embodied interactions for a holistic learning approach. Afterwards, we deploy the game as an optional learning intervention into a practical "SkillsLab" session of a bachelor's midwifery study program. Using the BYOD methodology, we investigate student acceptance, motivational benefits, and the games' usability.

II. Related Work

Investigating AR interventions for medical education through a systematic literature review in 2020, Tang et al. [4] found that the majority of available literature focuses on AR surgery training, followed by AR anatomy training. They not only concluded that coverage is sparse, but also, when assessing the quality of the evaluations, that current findings are insufficient to recommend adoption into educational curricula and more research is required.

A. AR for Anatomical Education

Furthermore, while there is no literature on specifically the female pelvis as the anatomy learning target and most focus on rather outdated AR approaches instead of modern techniques like handheld and head-mounted AR, secondary literature does point towards some primary studies on AR anatomical learning interventions in general, that are used as a basis for our work [5], [6].

Early explorations of AR as a learning intervention for anatomy-related topics were started in 2010. Here, Chien et al. [7] explored the anatomical structure of the skull, which could be decomposed and reassembled by medical students, using screen-based AR approaches. Thomas et al. [8] investigated the usage of a screen-based AR approach for the human ventricular system and found their system to be perceived as both useful and usable by the medical students that participated in the study. An AR textbook that uses screen-based AR for advanced visualizations of anatomical topics such as the lower limb system by Ferrer et al. [9] showed that AR can lead to better learning outcomes and higher test scores compared to a control group. In a comparative study, this AR intervention also outperformed conventional image-based and video learning methods in terms of acquired knowledge [10]. Ma et al. [11] explored the usability and user perception of a screen-based AR approach for bone, organ, and muscle visualizations, which were directly contextualized on the bodies of medical students using a webcam. In line with these efforts, other researchers also explored magic mirror approaches to contextualize anatomy content learning on the learners themselves. For example, Kugelmann et al. [12] used a magic mirror for gross anatomy learning in first-year medical students, which improved their perceived motivation and learning, and Bork et al. [13] and Moro et al. [14] explored their benefits and effectiveness for learning anatomy of organs.

In terms of handheld AR for anatomy learning, Jain et al. [15] explored the usage of tablets to visualize the anatomical structures of the human head on an AR marker. Similarly, Kurniawan et al. [16] used Smartphone-based AR with markers for the visualization of external and internal organs and found that it helped medical students to learn human anatomy. Jamali et al. [17] explored marker-based handheld AR as a learning intervention for the anatomy of the human skeletal structure and found that students were satisfied with their prototype in terms of its usability and set of features. Finally, Kücük et al. [18] showed that using handheld AR as an enhancement technique for a textbook, where figures can be enhanced to 3D AR animations through the smartphone, can increase academic achievements and lower cognitive load in second-year medical students.

B. Intrinsic Motivation

Games are commonly used in education as an approach to motivate students for the learning content. Research has shown that games can increase learning outcomes (e.g., [19]), students' motivation (e.g., [20]), and can foster long-term engagement through their motivational properties [21]. As such, games are usually designed to satisfy students' psychological needs for autonomy, competence, and relatedness. Hence, they can increase interest and engagement through their game mechanics and subsequently increase intrinsic motivation in general [21]. When students are intrinsically motivated, they feel interested, competent, and engage in an activity autonomously, leading to overall better performance. Extrinsic motivation, on the other hand, is controlled by outside factors (e.g., grades, rewards) and students are engaging in an activity rather through obligation or feeling pressured. This differentiation between extrinsic and intrinsic motivation is made through the self-determination theory [22]. In general, intrinsic motivation results in better learning outcomes and higher affection towards the learning experience compared to students who experience pressured learning environments, which undermine intrinsic motivation of students [23], [24]. Intrinsic motivation can also be fostered in game environments through providing game mechanics that support students' autonomy, competence, and relatedness and as such they can increase the engagement within the game environment. For the present study, we assess whether the developed AR game increases the intrinsic motivation of students.

III. The Pelvis AR Game

The Pelvis AR game that was developed for this study is a serious game for handheld Augmented Reality devices that allows students to interactively learn the regions and bone structures of the female pelvis. To ensure the game's scalability, no additional materials or AR-Markers are needed for its usage. During the game, students have to contextualize German-Latin word pairs to the correct region or bone structure of a 3D model of the female pelvis. To accomplish this task, they can either resolve the German-Latin pairing first and then combine the correct pair with the pelvis model, or combine the individual German and Latin pieces individually with the 3D pelvis model. The AR component of the game is used herein to enable embodied interactions, by requiring users to use deliberate hand and arm movements in physical space to pick up puzzle pieces and connect them with the correct area or part of the pelvis.

A. Flow of the game

After reading a case-based training description, the players complete the technical onboarding that explains the mechanics of AR and choose if they want to receive tips by an expert midwifery agent during the game. Then, they start the training by placing the virtual assembly with the pelvis and all puzzle pieces on a desk using markerless AR (see Section III-B and Fig 1 for the "design considerations" behind the purely virtual AR approach of the game). The game itself is split into 2 levels. In the first, the learning goal is familiarizing with the three bone regions and their German and Latin names. This is implemented as a labeling puzzle, where the two corresponding names (German and Latin) have to be identified within six pieces that are scattered around the pelvis. The players have to literally pick up the pieces by approaching them with the smartphone and grabbing them by the press of a button when they are in proximity to the center of the screen, which is signalled by a reticle. The grabbed piece then has to be overlapped with the matching piece or the matching region, and can be combined with both with an additional press of a button.

Afterwards, the more comprehensive task of contextualizing the 12 German and 12 Latin puzzle pieces for the bone-structures has to be completed using the same process. During the game, an optional agent, which serves as a virtual training partner, can provide textual and auditory tips to the user. If the app detects multiple or repeated incorrect interactions, an error overlay can provide feedback for actions that are not possible. After the game is completed, the app displays a summary and performance assessment to the user. Examples of the Pelvis AR game with a coarse training flow are visualized in Figure 1.

B. Design considerations

While the game is primarily targeted at midwifery bachelor students as a self-directed retention opportunity for a historically unpopular "dry" subject, it can also be used by nursing or medical students.

The idea behind it, in contrast to visualization-only approaches utilizing AR, is to use the purely virtual form of AR for contextualized visualization purposes and its interactivity to enable embodied interactions with the learning content by requiring the user to actively and deliberately utilize arm- and hand-movements in physical space to accomplish the tasks during the game. In contrast to conventional (non-AR) mobile serious games, in this, psychomotor learning is included into the learning task to provide a more holistic learning experience. Or, as Lindgren et al. [25] described it, "the design rationale is that having learners act out and physicalize the systems, processes, relationships, etc., that they are trying to understand [...] will create conceptual anchors from which new knowledge can be built." To incorporate affective learning considerations into the learning task, an optional midwifery agent can be activated that provides guidance and tips during the game. Therefore, the game incorporates all three domains of learning: cognitive, affective, and psychomotor. This should not only gamify the historically "dry" anatomy subject, but also increase intrinsic motivation for students to engage with it. There is also a good body of work indicating that increased interactivity of immersive technologies leads to improved learning outcomes and retention [1]–[3].

Additionally, the Pelvis AR game, with its purely virtual AR design, draws upon insights from Knierim et al. [26], who found that tangibility in AR trainings had no significant impact on learning outcomes and knowledge transfer, while significantly increasing setup-times. Therefore, while we still define it as AR, we utilize this very "VR-like" approach of purely virtual learning content in AR and still utilize Smartphones as the hardware choice without requiring additional physical material. Ultimately, this should increase the scalability of the AR intervention, enable Bring-Your-Own-Device approaches, and increase usability because of familiarity and introduction of very few obstacles for first-time users.

C. Technical development

From a technical perspective, the game was developed using the open-source TrainAR framework [27] in the Unity game engine and was deployed for Android and iOS for both mobile devices and tablets. It is published on the Apple App Store and the Google Play Store as part of the "Heb@AR App" [28], a free academic midwifery AR learning app that contains multiple AR trainings and games. In combination with the purely virtual AR approach, this allows the utilization of the Bring-Your-Own-Device methodology with its distinctive benefits [29] for this study and the published game beyond it.

IV. Study Design

The study was designed as an optional, non-controlled cohort within-subject before-and-after learning intervention. It was conducted during a practical "SkillsLab" tutorial session in an academic midwifery bachelor study program. A non-validated German translation of the Intrinsic Motivation Inventory [30], a validated German translation of the System Usability Scale (SUS) by Gao et al. [31] and qualitative feedback questionnaires were used in the pre- and post-study questionnaires to answer the following research questions:

1) Does the interactive handheld AR game increase the intrinsic motivation of students to engage in this anatomical terminology topic compared to their previous experiences with a conventional method?
2) Is the interactive handheld AR game usable by midwifery students?
3) Do students perceive the game as a valuable addition to existing learning methods? Could it potentially even replace the conventional methods?

A. Hypothesis

Our hypothesis is that the interactive properties of the Pelvis AR game increases students' intrinsic motivation to engage with the topic of anatomical terminology compared to

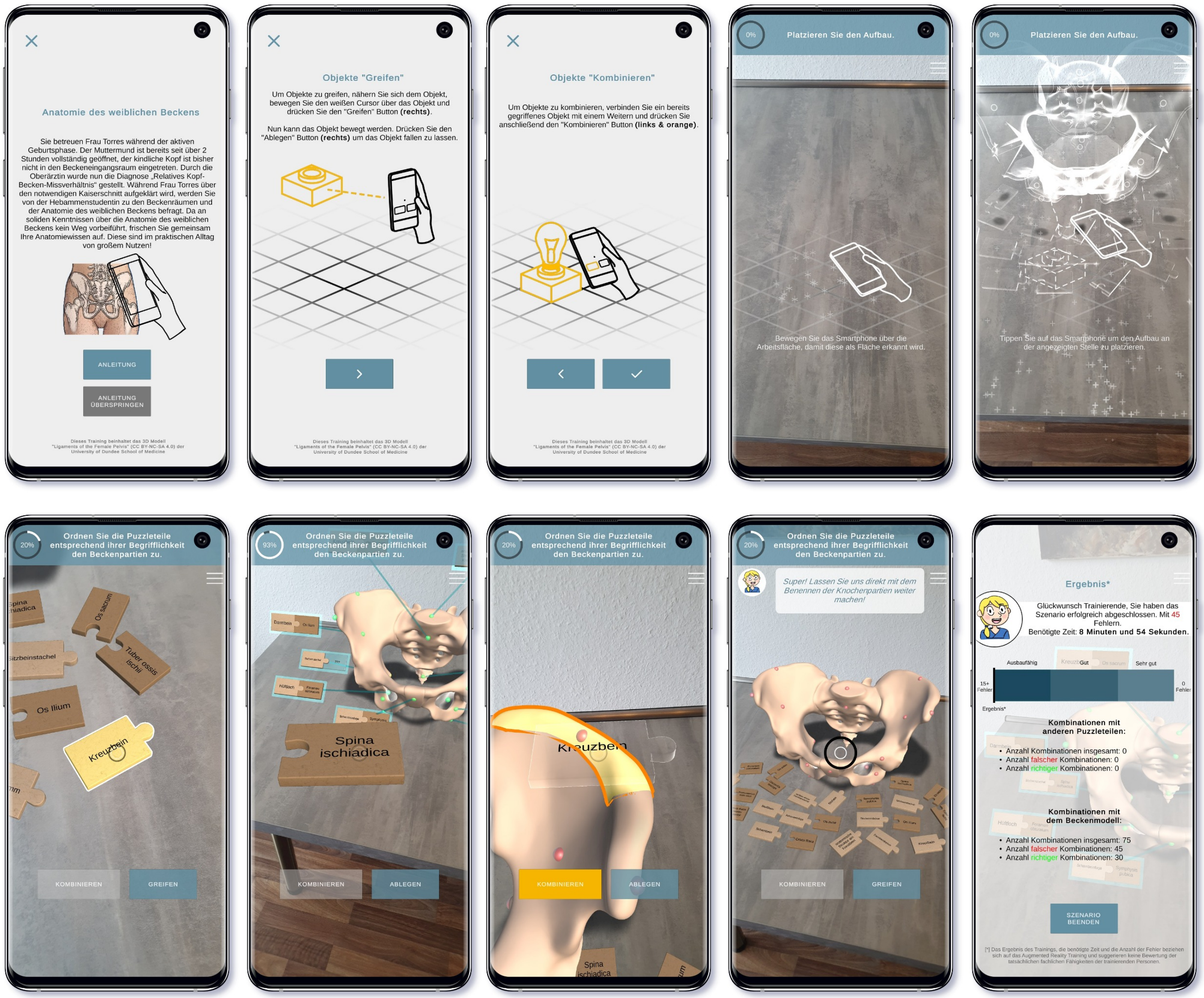

Fig. 1. The Pelvis AR serious game: Users are first shown a contextualized training case and are onboarded how to interact with the game. After scanning the environment and placing the Pelvis onto a table, they have to select, pick up and combine 30 puzzle pieces (3 Latin and 3 German for the pelvis regions, and 12 Latin and 12 German for the pelvis bone-structures) with the 3D Pelvis. An optional agent can provide additional textual/auditory tips during the game. After the game, users are shown a summary.

conventional methods significantly. To be more specific, this hypothesis can be split into the following three hypotheses: H1.1: the interactive Pelvis AR game increases perceived competence among students compared to previous experiences in a traditional memorizing exercise. H1.2: the interactive Pelvis AR game reduces the perceived pressure in studying content compared to previous experiences of the students. H1.3: the Pelvis AR game increases the intrinsic motivation of students compared to their previous experiences. Furthermore, as similar usability evaluations for other applications using the TrainAR framework had promising usability evaluations [27], [32], [33], we expect the usability of the interactive Pelvis AR game to be excellent for the target group and to not influence the motivational effects negatively by complicating the interaction unnecessarily (H2). When it comes to the students' acceptance, we expect them to perceive interactive handheld AR games as a useful optional addition to existing learning methods, but indicate skepticism regarding it potentially replacing conventional methods outright (H3).

B. Procedure

In a second semester practical midwifery lecture, in which students engaged with a handheld Augmented Reality procedural training for the first time, they were offered to participate in this study after completing the lecture's obligatory learning content. After scanning a QR code, participants first completed a pre-study survey on their smartphone. Here, they were asked for their consent, a demographic questionnaire, their

TABLE I
RESULTS OF THE THREE INDEPENDENT SAMPLES T-TEST EXAMINING THE CHANGE OF THE MOTIVATIONAL VARIABLES OF THE PERCEIVED COMPETENCE, PRESSURE/TENSION AND INTRINSIC MOTIVATION OF THE IMI [30] BEFORE AND AFTER PLAYING THE PELVIS AR GAME.

Parameter	**Pre-Test**		**Post-Test**		**t(31)**	**p**	**Cohen's d**
	M	SD	M	SD			
Perceived Competence	4.61	1.09	5.34	0.89	-4.56	<0.001	- 0.81
Pressure / Tension	2.83	1.35	3.43	0.57	-2.64	0.01	- 0.47
Intrinsic Motivation	4.60	1.22	5.50	0.63	-3.94	<0.001	- 0.67

experience with Latin and AR technology and the Intrinsic Motivation Inventory (IMI) [30] in relation to their conventional learning approach for anatomical terminology learning. Subsequently, students completed the AR Pelvis game, either through their own smartphones by downloading and then completing the game, or by using institutional smartphones that were made available to them. Finally, participants were asked to complete the post-study survey, where they answered the System Usability Scale, the Intrinsic Motivation Inventory on the Pelvis AR game, and qualitative questions on the perceived usefulness of the game. Participants completed the game independently and were not helped during the game. An experimenter was available in case of technical difficulties, and a midwifery professor was available in case of subject-related questions. Due to time and space constraints during the lecture, some participants had to complete the Pelvis AR game before filling out the pre-study survey.

V. RESULTS

The experiment was carried out with 36 primary qualifying midwifery bachelor students aged between 18 and 40, with an average age of 21.81 (SD = 4.52). All participants were female. Out of the 36 participants, 8 had the advanced Latin certificate (German: "Großes Latinum"), 4 had the intermediate Latin certificate (German: "Kleines Latinum") and the 24 remaining participants had no formal Latin certification. When asked how much experience they had with Augmented Reality, 33 participants reported having no experience with the technology, and 3 participants reported having very little experience.

While participation in the study was optional, it was attached to a practical "SkillsLab" training session of their curriculum, and students were not compensated for their participation. All participants in the study successfully completed the Pelvis AR learning game. 33 participants completed it on Apple iPhones, ranging from the iPhone 8, over the iPhone SE to the iPhone 12 Pro. Of the remaining participants, 2 used Android smartphones and 1 participant used an Android tablet to participate in the study.

A. Intrinsic Motivation

To measure the motivation of students, the three subscales interest, perceived competence, and pressure/tension of the IMI [30] were administered before and after the intervention. Hereby, 7-Point Likert Scales ranging from "strongly disagree" to "strongly agree" were utilized. For all three subscales, the Cronbach's alpha indicated sufficient internal consistency, ranging from α = 0.77 to α = 0.92. Additionally, for checking the differences between the pre- and post-measure, the assumption for normality was satisfied and therefore three paired t-tests were conducted using SPSS. To control for Type I error, a stricter p-value based on the Bonferroni correction (p = 0.017) was applied.

The results are displayed in Table I and show significant changes for all three subscales between the pre- and post-test. Perceived competence increased significantly after playing the Pelvis AR game compared to the students' previous experiences with traditional memorizing approaches t(31) = -4.56, p = <0.001. Moreover, also the perceived pressure/tension increased significantly after playing the Pelvis AR game t(31) = -2.64, p = 0.01. Finally, the result of the paired t-test shows that students' intrinsic motivation increased after the intervention (M = 5.50, SD = 0.63) compared to before the intervention (M = 4.60, SD = 1.22). This difference was highly significant t(31) = -3.94, p = <0.001. Both perceived competence (d = -0.81) and intrinsic motivation (d = -0.70) represented strong effect sizes, whereas the differences in pressure/tension represented a medium effect size of d = -0.47. The negative effect sizes indicate that the mean scores for all three subscales increased in the post measurement.

B. Perceived Usability

Calculating the results of the SUS [34] using the System Usability Scale Analysis Toolkit [38], revealed a SUS study score of 84.79 (SD = 13.51) with a minimum score of 55, a maximum score of 100 and a median score of 90. This SUS study score is considered an "Acceptable" usability according to Bangor et al. (2008) [34], would be graded an "A" on the empirical grading scale by Sauro et al. [35] and categorized as "Best Imaginable" usability when described using adjectives, according to Bangor et al. (2009) [36] (see Fig. 2, left). With a sample size of 36 participants, this result is 100% conclusive according to Tullis et al. [37] (see Fig. 2, upper right). Furthermore, when normalizing the 10 individual question scores (Question 1: "*I think that I would like to use this product frequently.*", Question 2: "*I found the product unnecessarily complex.*", Question 3: "*I thought this product was easy to use.*", Question 4: "*I think that I would need the support of a technical person to be able to use this product.*", Question 5: "*I found the various functions in this product were well integrated.*", Question 6: "*I thought there was too much inconsistency in this product.*", Question 7: "*I would imagine*

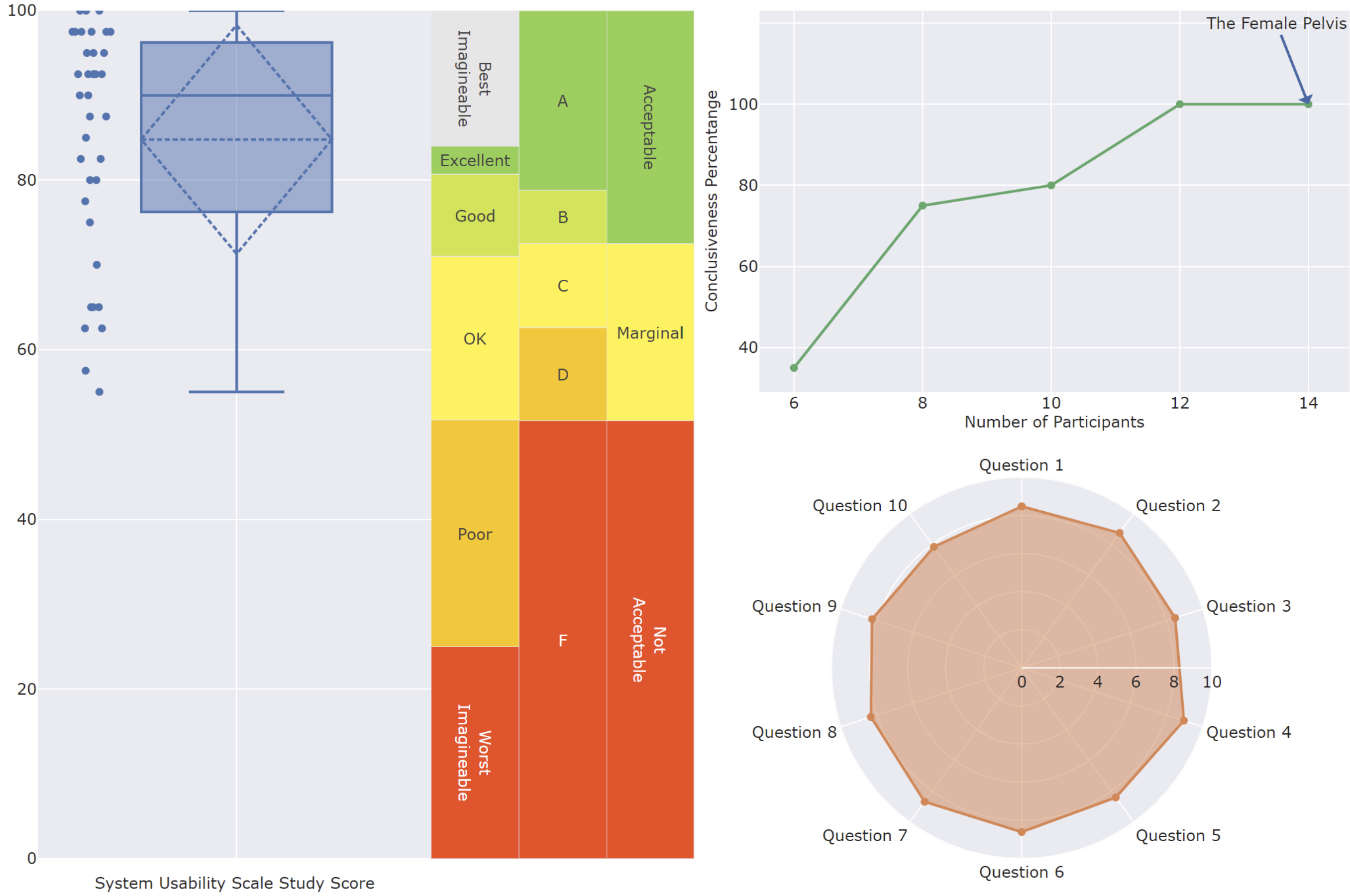

Fig. 2. The individual SUS scores as data points, the SUS study score as a box plot and the SUS contextualization scales: "Adjective Scale" [34], "Grade Scale" [35] and "Acceptability Scale" [36] for their interpretation (left). The conclusiveness percentage of the SUS study score based on the number of participants [37] (upper right) and the average contribution of each individual question of the SUS questionnaire [31] towards the SUS study score (lower right). Source: Plotted with the SUS Analysis Toolkit [38].

that most people would learn to use this product very quickly.", Question 8: "*I found this product very awkward to use.*", Question 9: "*I felt very confident using this product*", Question 10: "*I needed to learn a lot of things before I could get going with this product.*") to their average contribution towards the SUS study score according to Blattgerste et al. [38], there were no distinctive insights or deviations stemming from individual questions (see Fig. 2, lower right).

A one-way ANOVA revealed no statistical differences of perceived usability in form of SUS scores as a result of pre-existing latin certification, F (2, 33) = 1.915, p = 0.163.

C. Qualitative Feedback

Finally, to gather additional feedback on the perceived usefulness of the Pelvis AR game, we asked participants how much they agree with two statements on 7-Point Likert Scales and provide the reasoning for their answer.

When asked how much the participants agreed with the statement "I think the Pelvis AR training would help me learn the pelvic bones and pelvic spaces.", an average Likert scale value of 6.389 (SD = 1.231) was reported. This would indicate that the participants strongly agree with this statement. When asked for their reasoning, 14 participants provided no answer. Out of the remaining 22 participants, 8 indicated that they especially liked the 3D visualization as it would be particularly helpful for them to "envision the pelvis model spatially, as opposed to on paper". Additionally, 8 participants expressed their agreement with the concept of "active" learning in terms of naming and memorizing the parts and repeated correction in case of errors. One participant stated that the game would "make learning easier" and "have a long-term effect". Furthermore, three participants explained the perceived benefits mainly in connection with the fun and "playful" aspects of the game, which they expect to lead to "higher motivation" and the possibility of "using the game at home". Two participants stated that they had already acquired the knowledge through the conventional learning methods "books and index cards" or by using a "pelvis model" and the game would therefore not be helpful. Another participant stated that "the app is only useful with previous knowledge".

Subsequently, we asked how much participants agreed with the statement "I would rather use the Pelvis AR training instead of learning the pelvic bones and pelvic spaces the conventional way." and reported a Likert scale value of 5.222

(SD = 1.669), which would indicate that they agree with the statement. Sixteen participants provided no reasoning for their answer. Of the remaining participants, six noted that the advantage of training stems from the combination of realistic visualizations, which make it easier to "mentally associate and remember terms" compared to conventional learning methods. In this context, one of them pointed out that it is "less theoretical" (likely meaning less "dry" as a subject), which would "increase motivation" for her. Another participant mentioned that the advantage of this learning method is that it is "more convenient" when learning on the go. Two participants explained their reasoning for preferring the Pelvis AR game, with "it is more fun" and "exciting learning". The remaining participants were more critical. Two of them stated that they believe traditional learning methods to be "just as effective". In line with this perspective, five participants see "mixing" their traditional learning with the Pelvis AR game as promising so that the scenario rather serves "as a supplement" or complementary to "consolidate knowledge". One of the participants stated that although this learning method would not help her "learning the Latin terms", it would still help her in learning to contextualize the terminology onto the correct bones-structures and areas. Finally, two participants stated that they would prefer traditional learning methods because they did not appreciate "technical learning methods" and there is no possibility to "physically interact" with the Pelvis model.

Finally, when asked if participants had further feedback, notes or suggestions, three participants emphasized the meaningfulness of the "training" and thanked the developers. One participant noted that she had a lot of fun during the game.

VI. Discussion

While this study is exploratory in nature and part of a larger evaluation of the "Heb@AR App" [28] for midwives, several insights can be gathered from these results independently.

As expected, the Pelvis AR game increased the perceived competence (H1.1) and intrinsic motivation (H1.3) of the students compared to their previous experience with traditional methods. This is in line with previous research [1]–[3]. Perceived competence is theorized to support the overall intrinsic motivation [23], this could even elicit the measured result. Interestingly, hypothesis 1.2 had to be dismissed, since perceived pressure/tension increased after using the Pelvis AR game. There are multiple potential explanations for this result. On the one hand, this could have been a novelty effect, as the students were very inexperienced with AR in general and used the game for the first time. But arguably this should also have then been visible in the perceived usability, which was not evident in the results of the SUS. On the other hand, and the most likely explanation, this could have been caused by the game actually requiring one to contextualize all pieces successfully to complete the game's levels. This likely not only creates the immediate pressure of connecting every piece correctly to complete the game, but also the pressure of questioning how well the knowledge was actually previously acquired, as there was no validation method for the conventional learning method. There could also have been pressures created by this being part of a lecture with other students present, or the time constraints. Moreover, compared to the other two subscales, the difference in perceived pressure was smaller, although still significant. In general, the results indicate that the interactive handheld Pelvis AR Game increased the overall motivation of students through their perceived competence and intrinsic motivation scores significantly compared to the conventional learning methods they previously used (Q1).

The Pelvis AR game was not only usable by midwifery students, but achieved exceptionally high perceived usability scores compared to other immersive technologies and the highest recorded SUS study scores compared to other AR learning games and trainings using the TrainAR framework [27], [32], [33], answering Q2 with a clear yes. This is partially explained by the technical maturity of the framework and game at the time of this intervention but also by the fact that, compared to more complex procedural trainings with several interactions, the game could be completed with the same interaction of "combining" two objects that had to be repeated multiple times. Those usability scores were achieved despite the fact that the participants had virtually no previous AR experience. Additionally, pre-existing Latin certification, or the lack thereof, did not influence the usability of the Pelvis AR game. While this is to be expected, this indicates the learning content was successfully separated from the interaction metaphors of the game and the game is usable regardless of Latin knowledge, as desired.

Answering Q3, the qualitative feedback indicated that students perceived the interactive handheld AR game as a valuable addition to existing learning approaches, which was in line with our expectations (H3). When looking at the Likert scale value alone, participants seemed to be less critical of the Pelvis AR game that replaces conventional learning methods than expected. Looking at the qualitative feedback though, they did indicate that they would prefer a mix of both the conventional method and the Pelvis AR game, indicating that our statement might not have been clear enough and the results of the second qualitative statement have to be interpreted with caution. Notably, participants also focused on the "visualization" benefits of the game more often than on the "interactivity" and "fun" aspects in their qualitative feedback, which we perceived as the more important benefit in incorporating cognitive, affective, and psychomotor learning holistically. This, however, could also be a novelty effect as the visual aspects of AR are deviating the most from conventional methods and known forms of interactions with learning content, e.g., through smartphone apps or web-based trainings. Finally, while the engagement with the optional qualitative questions was relatively low, we believe this was no indication of little involvement in the study or disinterest in the Pelvis AR game, but rather caused by time-pressures during the testing in actual curricular usage and some participants filling out the questionnaire forms on Smartphones.

Combining the qualitative feedback on the perceived usefulness with the objective results of increased overall intrinsic motivation gathered through the IMI and contextualizing this with the perceived usability results of the SUS, we can discuss the results from the technology acceptance perspective. Although we do not directly apply the technology acceptance model by Davis et al. [39], it is clear that both usefulness and usability are satisfactory and students accept the Pelvis game as a learning intervention and, consequently, should form a behavioral intention to utilize it. This alone indicates that this interactive handheld AR game is an effective intervention, regardless of comparative learning benefits compared to conventional methods, as we expect it to motivate students to engage with the subject. While the AR learning benefits we found in this study are in line with previous work, this result adds to the literature in several ways: Scalable, cost-effective handheld AR devices can be utilized with BYOD approaches to achieve these results, optional anatomy learning interventions can be successfully applied in academic midwifery education and students can and want to independently use them, which should improve their self-determination.

Finally, the Pelvis AR game provides additional evidence, in line with previous work [33], that the TrainAR framework by Blattgerste et al. [27], originally envisioned as a framework for scalable procedural AR trainings, can also be applied to interactive non-procedural learning games as effective learning interventions.

A. *Limitations & Future Work*

This study provides valuable insights into the largely unexplored research field of motivational benefits of handheld AR applications for self-determined learning, and promising significant increases were found in terms of motivation in the study. However, the results are only a first exploration of the potential motivational benefits of AR games for anatomy learning and beyond. Thus, the results are considered preliminary, and we encourage replication. Generalizations should be made with caution and larger scale studies are needed, especially comparing these effects with a control group.

Moreover, while we argue that motivational benefits alone justify the usefulness and effectiveness of the learning intervention, showing that the Pelvis game is more effective and efficient in terms of learning outcomes in comparative studies would likely strengthen the contribution and should be investigated in future work. While, based on previous work, we theorize and strongly expect comparative learning benefits from the interactivity of the AR game based on the more holistic learning approach, this still has to be shown empirically. We support large-scale studies from educators on this topic by making the game publicly available as part of the "Heb@AR App" [28] in all relevant app stores.

Furthermore, while we were able to show motivational benefits when applying the intervention to first-time users, there could be an uncontrolled novelty effect on motivation, and it could decrease with familiarity during actual curricular usage over time [3].

As the results are promising, and the limitations largely require longer-term actual usage data, we are currently working on sophisticated tracking and in-app self-reporting mechanisms that can be integrated into the game to analyze learning benefits, progress, and actual usage frequency to get further insights.

VII. Conclusion

In this work, the Pelvis AR game is presented and evaluated with its design considerations. It is a scalable, interactive handheld AR game for midwifery students that allows self-determined learning of the bone structures and areas of the female pelvis and is available as part of a midwifery training app in the Android and iOS App Stores, called the "Heb@AR App" [28]. The game was evaluated in terms of its motivational benefits, perceived usability, and perceived usefulness. Participants of the voluntary learning intervention showed increased perceived competence, intrinsic motivation, but also increased pressure/tension after playing the game. Furthermore, they indicated that the game had exceptional usability and stated that they perceived the game as a useful complementary learning tool.

While designed and evaluated for the specific learning intervention of the female pelvis, the design considerations and insights should be applicable to a broad range of anatomy learning topics in the medical context, and beyond that to other vocabulary or technical terminology learning contexts.

Acknowledgment

The authors would like to thank Nicola H. Bauer and Kristina Vogel for providing the "female pelvis" as the chosen learning intervention, Annette Bernloehr for her support in the evaluation, and the midwifery students at Bielefeld University of Applied Sciences for their participation in the experiment. This research was supported by the grant 16DHB3021, project "HebAR - AR-Based-Training-Technology", by the German Ministry for Education and Research (BMBF). Michaela Arztmann has been supported by funding from the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement No 812716.

References

[1] J. Garzón, J. Pavón, and S. Baldiris, "Systematic review and meta-analysis of augmented reality in educational settings," *Virtual Reality*, vol. 23, no. 4, pp. 447–459, 2019.

[2] N. M. Alzahrani, "Augmented reality: A systematic review of its benefits and challenges in e-learning contexts," *Applied Sciences*, vol. 10, no. 16, p. 5660, 2020.

[3] M. Akçayır and G. Akçayır, "Advantages and challenges associated with augmented reality for education: A systematic review of the literature," *Educational Research Review*, vol. 20, pp. 1–11, 2017.

[4] K. S. Tang, D. L. Cheng, E. Mi, and P. B. Greenberg, "Augmented reality in medical education: a systematic review," *Canadian medical education journal*, vol. 11, no. 1, p. e81, 2020.

[5] D. Chytas, E. O. Johnson, M. Piagkou, A. Mazarakis, G. C. Babis, E. Chronopoulos, V. S. Nikolaou, N. Lazaridis, and K. Natsis, "The role of augmented reality in anatomical education: An overview," *Annals of Anatomy-Anatomischer Anzeiger*, vol. 229, p. 151463, 2020.

[6] M. Duarte, L. Santos, J. G. Júnior, and M. Peccin, "Learning anatomy by virtual reality and augmented reality. a scope review," *Morphologie*, vol. 104, no. 347, pp. 254–266, 2020.

[7] C.-H. Chien, C.-H. Chen, and T.-S. Jeng, "An interactive augmented reality system for learning anatomy structure," in *proceedings of the international multiconference of engineers and computer scientists*, vol. 1. International Association of Engineers Hong Kong, China, 2010, pp. 17–19.

[8] R. G. Thomas, N. William John, and J. M. Delieu, "Augmented reality for anatomical education," *Journal of visual communication in medicine*, vol. 33, no. 1, pp. 6–15, 2010.

[9] J. Ferrer-Torregrosa, J. Torralba, M. Jimenez, S. García, and J. Barcia, "Arbook: Development and assessment of a tool based on augmented reality for anatomy," *Journal of Science Education and Technology*, vol. 24, no. 1, pp. 119–124, 2015.

[10] J. Ferrer-Torregrosa, M. Á. Jiménez-Rodríguez, J. Torralba-Estelles, F. Garzón-Farinós, M. Pérez-Bermejo, and N. Fernández-Ehrling, "Distance learning ects and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes," *BMC medical education*, vol. 16, no. 1, pp. 1–9, 2016.

[11] M. Ma, P. Fallavollita, I. Seelbach, A. M. Von Der Heide, E. Euler, J. Waschke, and N. Navab, "Personalized augmented reality for anatomy education," *Clinical Anatomy*, vol. 29, no. 4, pp. 446–453, 2016.

[12] D. Kugelmann, L. Stratmann, N. Nühlen, F. Bork, S. Hoffmann, G. Samarbarksh, A. Pferschy, A. M. Von der Heide, A. Eimannsberger, P. Fallavollita *et al.*, "An augmented reality magic mirror as additive teaching device for gross anatomy," *Annals of Anatomy-Anatomischer Anzeiger*, vol. 215, pp. 71–77, 2018.

[13] F. Bork, R. Barmaki, U. Eck, K. Yu, C. Sandor, and N. Navab, "Empirical study of non-reversing magic mirrors for augmented reality anatomy learning," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2017, pp. 169–176.

[14] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, "The effectiveness of virtual and augmented reality in health sciences and medical anatomy," *Anatomical sciences education*, vol. 10, no. 6, pp. 549–559, 2017.

[15] N. Jain, P. Youngblood, M. Hasel, and S. Srivastava, "An augmented reality tool for learning spatial anatomy on mobile devices," *Clinical Anatomy*, vol. 30, no. 6, pp. 736–741, 2017.

[16] M. H. Kurniawan, G. Witjaksono *et al.*, "Human anatomy learning systems using augmented reality on mobile application," *Procedia Computer Science*, vol. 135, pp. 80–88, 2018.

[17] S. S. Jamali, M. F. Shiratuddin, K. W. Wong, and C. L. Oskam, "Utilising mobile-augmented reality for learning human anatomy," *Procedia-Social and Behavioral Sciences*, vol. 197, pp. 659–668, 2015.

[18] S. Küçük, S. Kapakin, and Y. Göktaş, "Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load," *Anatomical sciences education*, vol. 9, no. 5, pp. 411–421, 2016.

[19] P. Wouters, C. van Nimwegen, H. van Oostendorp, and E. van der Spek, "A meta-analysis of the cognitive and motivational effects of serious games," *Journal of Educational Psychology*, vol. 105, pp. 249–265, 2013.

[20] R. L. Lamb, L. Annetta, J. Firestone, and E. Etopio, "A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations," *Computers in Human Behavior*, vol. 80, pp. 158–167, 2018.

[21] R. M. Ryan and S. C. Rigby, "Motivational foundations of game-based learning," in *Handbook of Game-Based Learning*, . B. H. J.L. Plass, R.E. Mayer, Ed. Cambridge: The MIT Press, 2020, pp. 153–176.

[22] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist*, vol. 55, no. 1, pp. 68–78, 2000.

[23] ——, "Self-determination theory," in *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, 2nd ed., J. D. Wright, Ed. Oxford: Elsevier, 2015, pp. 486–491.

[24] E. L. Deci and R. M. Ryan, *Optimizing Students' Motivation in the Era of Testing and Pressure A Self-Determination Theory Perspective*. Singapore: Springer Singapore, 2016, pp. 9–29.

[25] R. Lindgren, M. Tscholl, S. Wang, and E. Johnson, "Enhancing learning and engagement through embodied interaction within a mixed reality simulation," *Computers & Education*, vol. 95, pp. 174–187, 2016.

[26] P. Knierim, F. Kiss, M. Rauh, and A. Schmidt, "Tangibility is overrated: Comparing learning experiences of physical setups and their virtual equivalent in augmented reality," in *19th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 299–305. [Online]. Available: https://doi.org/10.1145/3428361.3428379

[27] J. Blattgerste, K. Luksch, C. Lewa, and T. Pfeiffer, "TrainAR: A scalable interaction concept and didactic framework for procedural trainings using handheld augmented reality," *Multimodal Technologies and Interaction*, vol. 5, no. 7, p. 30, 2021.

[28] J. Blattgerste, C. Lewa, K. Vogel, T. Willmeroth, S. Janßen, J. Franssen, J. Behrends, M. Joswig, T. Schäfer, N. H. Bauer, A. Bernloehr, and T. Pfeiffer, "Die Heb@AR App - eine android & iOS app mit augmented reality trainings für selbstbestimmtes und curriculares lernen in der hochschulischen hebammenausbildung," in *Wettbewerbsband AVRiL 2022*, H. Söbke and R. Zender, Eds. Bonn: Gesellschaft für Informatik e.V., 2022, pp. 4–9. [Online]. Available: https://doi.org/10.18420/avril2022_01

[29] S. Pozo Sanchez, J. López-Belmonte, A.-J. Moreno-Guerrero, J. M. Sola Reche, and A. Fuentes Cabrera, "Effect of bring-your-own-device program on flipped learning in higher education students," *Sustainability*, vol. 12, no. 9, p. 3729, 2020.

[30] R. M. Ryan, "Intrinsic motivation inventory (imi)," 1982, accessed: 2022-08-15. [Online]. Available: https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

[31] M. Gao, P. Kortum, and F. L. Oswald, "Multi-language toolkit for the system usability scale," *International Journal of Human–Computer Interaction*, vol. 36, no. 20, pp. 1883–1901, 2020.

[32] J. L. Dominguez Alfaro, S. Gantois, J. Blattgerste, R. De Croon, K. Verbert, T. Pfeiffer, and P. Van Puyvelde, "Mobile augmented reality laboratory for learning acid–base titration," *Journal of Chemical Education*, vol. 99, no. 2, 2022.

[33] M. Arztmann, J. L. D. Alfaro, J. Blattgerste, J. Jeuring, and P. van Puyvelde, "Marie's ChemLab: a mobile augmented reality game to teach basic chemistry to children," in *European Conference on Games Based Learning*, ser. ECGBL '22, 2022.

[34] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.

[35] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[36] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[37] T. Tullis and J. Stetson, "A comparison of questionnaires for assessing website usability," *Usability professional association conference 1*, 06 2006.

[38] J. Blattgerste, J. Behrends, and T. Pfeiffer, "A web-based analysis toolkit for the system usability scale," in *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 237–246. [Online]. Available: https://doi.org/10.1145/3529190.3529216

[39] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.