JMIR Publications
The Leading eHealth Publisher

Applying Augmented Reality to Convey Medical Knowledge On Osteoclasts: A User Study

Jascha Grübel, Julia Chatain, Claudio Schmid, Violaine Fayolle, Fabio Zünd, Reinhard Gruber, Bernd Stadlinger

Submitted to: JMIR Serious Games
on: July 30, 2024

Table of Contents

Applying Augmented Reality to Convey Medical Knowledge On Osteoclasts: A User Study

Jascha Grübel[1, 2, 3, 4, 5] Dr. sc. ETH Zurich; Julia Chatain[6, 3] Dr. sc. ETH Zurich; Claudio Schmid[7] Dr. med. dent.; Violaine Fayolle[3, 8] MA; Fabio Zünd[3] Dr. sc. ETH Zurich; Reinhard Gruber[9] Dr. nat. techn.; Bernd Stadlinger[7] Dr. med., Dr. med. dent.

[1]Laboratory for Geo-Information Science and Remote Sensing Environmental Sciences Group Wageningen University Wageningen NL
[2]Center for Sustainable Future Mobility Dept. of Mechanical and Process Engineering & Dept. of Civil, Environmental and Geomatic Engineering ETH Zurich Zürich CH
[3]Game Technology Center Department of Computer Science ETH Zurich Zürich CH
[4]Chair of Cognitive Science Department of Humanities, Social and Political Sciences ETH Zurich Zürich CH
[5]Visual Computing Group Harvard John A. Paulson School Of Engineering And Applied Sciences Harvard University Boston US
[6]Future Embodied Learning Technologies Singapore-ETH Centre Singapore SG
[7]Center for Dental Medicine University of Zurich Zürich CH
[8]Disney Research Zürich CH
[9]Competence Center of Oral Biology University Clinic of Dentistry Medical University of Vienna Vienna AT

Corresponding Author:
Jascha Grübel Dr. sc. ETH Zurich
Laboratory for Geo-Information Science and Remote Sensing
Environmental Sciences Group
Wageningen University
Droevendaalsesteeg 4
Wageningen
NL

Abstract

Background: Augmented Reality (AR) as an interactive communication tool has matured to the point that it can be used in the classroom to engage medical students.

Objective: We assess a state-of-the-art AR game published together with a modern cell atlas on cells of the oral cavity to investigate the trade-offs of the new technology to convey medical knowledge.

Methods: A serious game is designed and created to represent state-of-the-art knowledge on osteoclasts for the classroom. The game is evaluated for its usability and a vignette experiment is conducted comparing the topic of Osteoclasts in the AR game and a textbook option conveying the same information. Participants are randomly assigned and their learning success is measured after the treatment, one week later and one month later. We also assess the use of the game in the app store to gain general insights.

Results: A serious game elicits strong interest in the topic and motivates students. The learning outcomes are comparable to text-based self-learning but with higher engagement. Furthermore, curious students benefit more from interactive learning methods compared to text-only methods and have higher learning success.

Conclusions: Introducing new technology like AR into teaching curricula requires technological investment, updated curricula and careful application of learning paradigms. AR-based learning may serve especially curious students that usually learn less in text-heavy teaching. In addition, a new generation of students is soon entering the classroom that is very used to short and entertaining information exchanges and our methods repertoire needs to adapt to these new realities.

(JMIR Preprints 30/07/2024:64751)
DOI: https://doi.org/10.2196/preprints.64751

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Applying Augmented Reality to Convey Medical Knowledge on Osteoclasts: A User Study

Figure 1: The AR Osteoclast serious game being played in a lecture hall.

Abstract

Background: Augmented Reality (AR) as an interactive communication tool has matured to the point that it can be used in the classroom to engage medical students.

Objective: We assess a state-of-the-art AR game published together with a modern cell atlas on cells of the oral cavity to investigate the trade-offs of the new technology to convey medical knowledge.

Methods: A serious game is designed and created to represent state-of-the-art knowledge on osteoclasts for the classroom. The game is evaluated for its usability and a vignette experiment is conducted comparing the topic of Osteoclasts in the AR game and a textbook option conveying the same information. Participants are randomly assigned, and their learning success is measured after the treatment, one week later and one month later. We also assess the use of the game in the app store to gain general insights.

Results: A serious game elicits strong interest in the topic and motivates students. The learning outcomes are comparable to text-based self-learning but with higher engagement. Furthermore, curious students benefit more from interactive learning methods compared to text-only methods and have higher learning success.

Conclusions: Introducing new technology like AR into teaching curricula requires technological investment, updated curricula and careful application of learning paradigms. AR-based learning may

serve especially curious students that usually learn less in text-heavy teaching. In addition, a new generation of students is soon entering the classroom that is very used to short and entertaining information exchanges and our methods repertoire needs to adapt to these new realities.

Keywords: Tablet-based Augmented Reality; Osteoclasts; Educational Game; User Study; Vignette Experiment

Introduction

Active learning strategies focus on the students' autonomy and agency over their learning process [1]. The maturity of visualization technology opens new doors for interactive learning paradigms [2]. The application of immersive media [3] with Augmented Reality (AR) is an interesting option to enrich medical teaching (Figure 1). However, the study of the educational benefit of interactive content is still in its infancy. One potential pathway for improving the learning experience could be a reduced cognitive load through the visualizations [4]. It is important to both match content and medium to produce a coherent learning experience for the students [2]. In this study, we explore both the content generation for immersive learning and the impact it has on student performance.

Knowledge of physiology is essential for teaching health sciences, including medicine and dentistry. For surgical subjects such as orthopedics or oral & maxillofacial surgery, bone physiology is of special interest [5]. Many therapies including dental implant placement are based on the success of bone remodeling [6]. Therefore, the understanding of bone biology in general and bone remodeling is important. In medical school, this is currently taught mainly in histology, where these tissues are visualized and discussed at the microscopic level [7]. Nowadays histology classes include lectures and some microscopy lab components. The spatial understanding of two-dimensional (2D) sections is intrinsically challenging for the students [8].

In the past decades, active learning strategies and tools have been developed to improve teaching (i.e., flipped classroom, problem-based learning, productive failure). Many studies have already shown that such methods can improve understanding, long-term learning, transfer, and the development of self-directed learning skills [9–12]. In this context, serious gaming for teaching is interesting. A study performed by Felszeghy et al. [13] showed that the use of gamification in histology teaching can enhance student learning.

Augmented reality (AR) is a relatively new technology that enables the integration of real environmental stimuli into a digitally generated three-dimensional (3d) representation [3,14]. It is a type of technology in which computer-generated virtual content is superimposed over real structure, enhancing the sensory perception of reality [15]. From a learning stand-point, AR is particularly interesting as it enables active forms of learning [16], and co-locates information and its referent, thus reducing split-attention effects [17], while also supporting collaboration and social incentives to learning [18]. This raises the question, whether a hands-on immersive AR experience may help to understand complex physiological processes.

In the early 1990s first basic computer-assisted anatomy programs appeared [14]. The continuing development of hard- and software technologies and the availability of modern head-mounted displays (e.g. Microsoft HoloLens) have opened new possibilities for teaching and learning tools

[19]. AR has also benefited from the continuous improvement of smartphones and tablet devices [20]. Examples for the use of AR are the Augmented Creativity Coloring Book app [21], which brings colored drawings to life in AR, and the Viseum app [22], which "augments" an artwork in an art exhibition with contextualized information fostering art education. Both apps were designed under the auspices of the Game Technology Center, ETH Zurich, Switzerland. They present 3D objects, images, videos, or audio superimposed over the real world and enable an interaction in a mixed reality space. In general, AR starts to influence everyday life; to name a few examples: Snapchat filters, Pokémon Go, IKEA Place [23]. But medical education has also seen applications, e.g. in nursing education [24] and has been generally considered useful [25].

The aim of this project was to "augment" a medical textbook / cell atlas with a strong focus on photorealistic imaging (Visual Biology in Oral Medicine, Quintessence Publishing [5]) by means of an AR-based App. The text book itself was partially inspired by six computer-animated scientific videos on cell communication processes in a clinical context and state of the art reviews [24–27]. The App *AR Osteoclasts* was designed to enable an enhanced visual experience of bone physiology, see Figure 1. The process of bone resorption and new bone formation is described in the underlying book chapter on osteoclasts (p. 145. The AR-App enables the user to experience these processes of bone remodeling actively by interaction with an AR-world. With the present study we analyze whether dental students show an improved understanding in bone physiology, after having played this App. Further, the level of motivation was assessed.

We formulate the following hypotheses based on previous learning literature and the topic of osteoclasts:

H1a: Learning about Osteoclasts with a serious AR game is as least as effective as learning from text.

H1b: The AR experience results in better learning outcomes on visual questions.

H2: Learning in AR is beneficial for curious learners, who benefit more from active learning as they tend to explore more.

H3: Topics around the visual components of the game are more easily learned from the AR experience.

H4: Long-term learning is improved through experiential learning in the AR game.

Methods

This paper relies on two kinds of methodology to assess Augmented Reality Serious Gaming for Learning. First, we present our approach for the usability study focused on gaining insights into the effectiveness of the present game. Second, we present a learning experiment, where we explore how AR learning compares to classical text learning. We use a difference-in-differences design to assess learning outcomes quantitatively and automated text-analysis to assess the outcomes qualitatively. In the Appendix D, we present how the game was designed to support the mechanisms for learning with AR.

Usability study

The usability study has two goals, first to validate that users can experience the full content of the game. Second, to understand their subjective experience of the game mechanisms. For an optimal learning experience, we want to ensure that users understand the game mechanisms so that we do not test their skills in operating the game instead of learning outcomes [28].

Recruitment

We investigate the usability of AR osteoclasts from two sources. First, the game has been made available with a QR code in the book. This brings in expert users and students to play the game as they use the book. The engagement with the game is optional. We engage these with a voluntary survey after playing the game that resembles the survey from the student learning experiment. Participants gave informed consent before starting the survey. Second, we use our learning experiment to also evaluate the usability in a controlled context (Appendix 1A.1 Figure S1). The recruitment for in-person participants is detailed in the section on the student learning experiment.

Methods

The survey after the AR experience evaluates the design of our game in terms of user experience. For the course students, we employ the 1short version of the User Experience Questionnaire (UEQ) [29,30], that is the UEQ-S [31], to increase retention [32]. These questionnaires allow a distinction between objective (pragmatic) and subjective (hedonic) qualities of the experience. To understand the role of AR in the resulting experience, engagement, and learning outcomes, we investigate the immersion using the Augmented Reality Immersion (ARI) questionnaire [33]. This is crucial to understand whether students were indeed immersed in the activity and could benefit fully from the AR capabilities.

Finally, we investigate how well participants were able to use the game by looking at several key indicators. First, playtime is gathered by stages to assess how long people engaged with the mini games as well as with the encyclopedia. Second, activation of the AR scene is counted to check how well people could operate and interact with the AR scene. As we use marker-based AR, the scene is activated when the picture presented in the book is sufficiently in the tablet's field of view and lost otherwise. Lastly, we assess the user actions in the AR game as either being a “success”, a “miss”, or a “fail”. A success is an interaction that was performed according to the game design. A miss is an interaction that is close to the game design but does not trigger the desired outcome. For example, in the first part of Mini-Game 1, a circle is drawn not on the bone surface but the blood vessel. Lastly, a “fail” is an interaction that was not planned in the game design. It is noteworthy that explorative behavior may be considered a failure in this setup.

Student Learning Experiment

Recruitment

The dental students of the last year of dental school attending a course at the end of their studies, being adequate to the topic and content level were selected. All students were informed in advance about the study taking place during class. The study was explained to the students before the start. Participation was optional. Students gave informed consent before participating. The study was approved by the ethics commission of the University of Zürich (2024-258707).

Methods

As opposed to linear expository texts, our serious game leverages students' curiosity and exploration affinity, which, in turn, may impact learning outcomes. To account for this, we employ the curiosity and exploratory inventory II (CEI-II) [34], focusing on curiosity trait (general curiosity), as we are interested in how students' curiosity trait impacts their experience, as opposed to how the experience impacts students' curiosity state (curiosity at the time of playing).

We used three types of learning constructs to evaluate the students' learning outcomes. First, we implemented questions that require recall of knowledge attained in the game. This is the simplest form of knowledge retention. Second, we implemented questions that require transfer of knowledge to apply what they learned in the game to new problems or contexts [35]. Lastly, we augmented knowledge transfer with visual knowledge transfer where students are presented with a visualization of a process and are asked to transfer knowledge from their game experience. This is important as transferring to other representations requires both an understanding of the content and of the representation (representation dilemma), making it a more difficult transfer task [36]. The visualizations are not produced in-game and are unrelated scientific visualizations.

Human motivation is often characterized by the ARCS (Attention, Relevance, Confidence, and Satisfaction) model [37,38] which in turn is solicited with the Instructional Materials Motivation Survey (IMMS). Attention refers to how well the learning materials catch the attention and interest of the learner. Relevance relates to how relevant and connected to the real world the materials are perceived to be by the learners. Confidence captures how confident the learners feel regarding their future success in the discipline when using the materials. Lastly, Satisfaction defines how satisfied the learners are with their own progress when learning with the materials.

The students are randomly assigned into two separate treatment groups where one receives only a text describing bone remodeling and the other receives first the same text and then is allowed to play the game (Figure 2). The text describes the processes that take place in the game and thus both can be considered to provide the same information. To better assess knowledge retention, the questionnaires on the knowledge attainment are presented to the participants three times. First, they are presented directly after the learning experience. Second, they are presented one week after the learning experience and lastly, they are presented one month after the learning experience. The usability study is conducted only with the participants of the AR game.

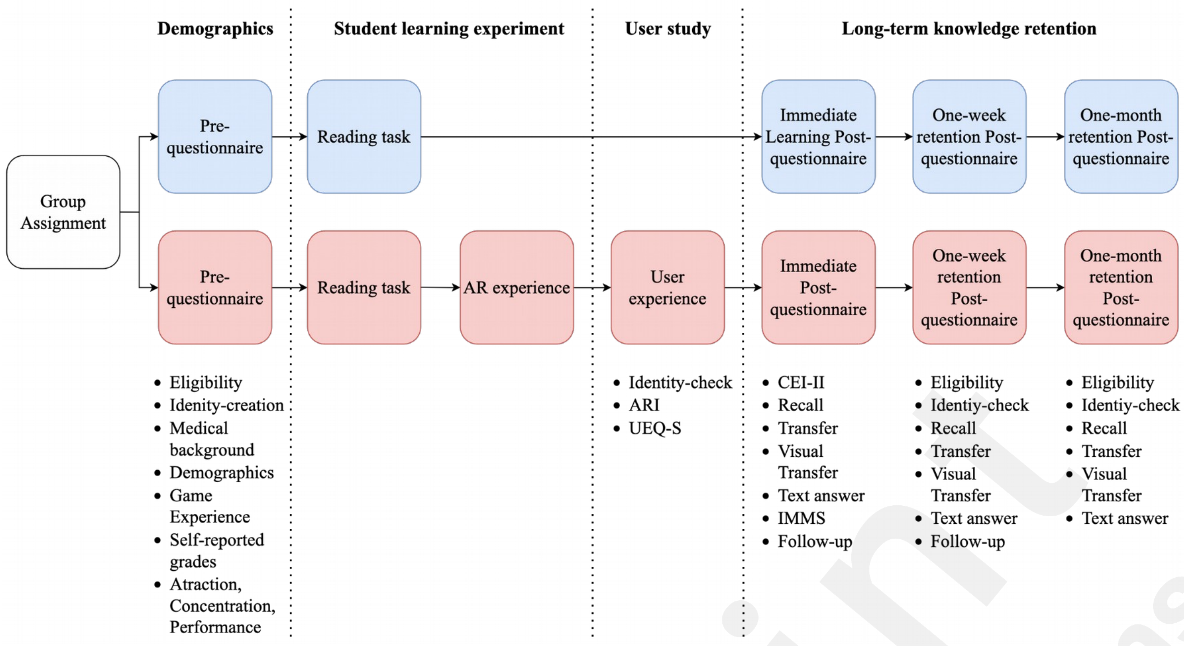


Figure 2: Protocol for the student learning experiment. Students are randomly assigned into the text-only group (blue) and the AR experience (red). General demographic information is acquired before the experiment and user study and the outcome is measured three times to gauge long-term retention of knowledge.

Statistical Analysis

Power

Given that we are analyzing the difference for a binary classification, we can take the difference in recall rate (remember an answer correctly) for each treatment as our variable of interest. We aim for an effect size of 10% difference in the recall rate with a standard deviation of 10% in recall rate. At an alpha level of 0.05 and power of 0.8, this amounts to a total of 32 required participants (at least 16 per group).

Data Exclusion

No data points from the 39 participants were excluded by principle. However, attrition removed 2 participants during the experiment. The first follow-up survey to recall the learned material only had 6 responses and the second follow-up survey to recall the learned material only had 15. The follow–up studies do not allow for statistical evaluation given our power analysis.

Linear regression model

We develop linear regression models for our three learning constructs, Recall, Transfer and Visual Transfer as well as a composed model that joins the performance across all three constructs. Each model has the learning construct as the dependent variable and the following independent variables:

$$y=\beta_0+\beta_1\cdot Treatment+\beta_2\cdot CEI2+\beta_3\cdot ARI+\beta_4\cdot Grade+\beta_5\cdot Age+¿$$

$$\beta_6\cdot Gender+\beta_7\cdot Experience+\beta_8\cdot Image+\beta_9\cdot Interaction+¿$$

$$\beta_{10} \cdot Text + \beta_{11} \cdot Treatment \cdot CEI2 + \varepsilon$$

The *Treatment* is defined as either being AR or Text-only learning. The *CEI2* is the sum across all CEI-II items. The *ARI* is the sum across all ARI items. The *Grade* is the Swiss grade as reported by the student, between 1 (worst) and 6 (best). The *Age* is a self-reported 5-year-age bracket converted to a continuous scale. The *Gender* is either male, female or other. In our sample, we did not have the latter category. The *Experience* refers to the expertise participants had with osteoclasts before scored as a 5-point scale from no knowledge to regular interaction. The *Image*, *Interaction* and *Text* refer to preferred modes of learning as self-reported by the student. The β_{11} is the coefficient for the interaction between the treatment group and the CEI-II.

As interaction models are notoriously difficult to interpret [39], we use marginal effects to understand the impact that the interaction between treatment and CEI-II has. We use unit-level conditional estimates of the empirical distribution to compute the marginal effects for the full range of the CEI-II scale from 10 to 50 points with heteroskedasticity-robust standard errors at the 90% confidence interval [40].

Text Analysis

Students were asked to answer three text-based questions about bone remodeling (Table 2). We used encouragement designs [41] to ask for a longer answer. The text answers provided by students are analyzed with structural topic modeling [29] to understand the relationship between the students' responses and the treatments. Topic modeling in general finds themes in a large set of written documents that are coherent. Structural topic modeling expands these tools with the option to add covariates to the documents that contain meta information not included in the documents themselves. We use structural topic modeling to annotate the treatment, which question they answered, which texts come from the same person, and how many weeks after the treatment they answered the question.

Table 2: The questions for the text analysis.

	Title	Question
Q1	Formation	*Describe the formation of an osteoclast in 3 to 6 sentences.*
Q2	Function	*Describe in words the function of osteoclasts and osteoblasts in the canopy in 3 to 6 sentences.*
Q3	Differentiation	*Describe the processes and molecules involved in osteogenic differentiation, and the role they play in 3 to 6 sentences.*

Topic modeling is a statistical iterative process and largely depends on initialization [29]. Furthermore, the number of topics is *a priori* unknown and needs to be determined. First, we run a large set of models with different numbers of topics from 2 to 20 multiple times. We also investigate the behavior for different numbers of topics with and without covariates (see appendix A1). We determine that K=10 topics is optimal for the text data we acquired. To counter issues of initialization, we run 100 models with K=10 topics and choose the best model for the trade-off of exclusivity and coherence.

The topics are labeled by three independent experts and the intercoder reliability index Krippendorff's α [42] is calculated to determine the quality of our labels. Krippendorff's α ranges from -1 to 1 with 1 indicates a perfect match, 0 indicates randomly chosen labels and -1 indicates inverse labeling [43]. Typically, ratings of 0.8>α indicate high reliability and 0.67>α indicate acceptable reliability [42].

Survey Analysis

To understand how participants subjectively perceive the learning in AR, we calculate the differences for responses to the ARCS model with a Wilcoxon signed rank test [44]. The full IMMS questionnaire [38] is available in the appendix (Multimedia Appendix 2F).

Results

First, we present the results of the usability study to understand how well people take on the AR experience and whether they were able to learn with the game. Second, we explore the effectiveness of our difference-in-differences design for comparing text-based learning with AR-based learning.

Usability study

Surveys

We assessed the Augmented Reality Immersion (ARI) of participants and found that generally people online were more immersed. To some degree, this may be due to the classroom setting of the student experiment (Figure 3).

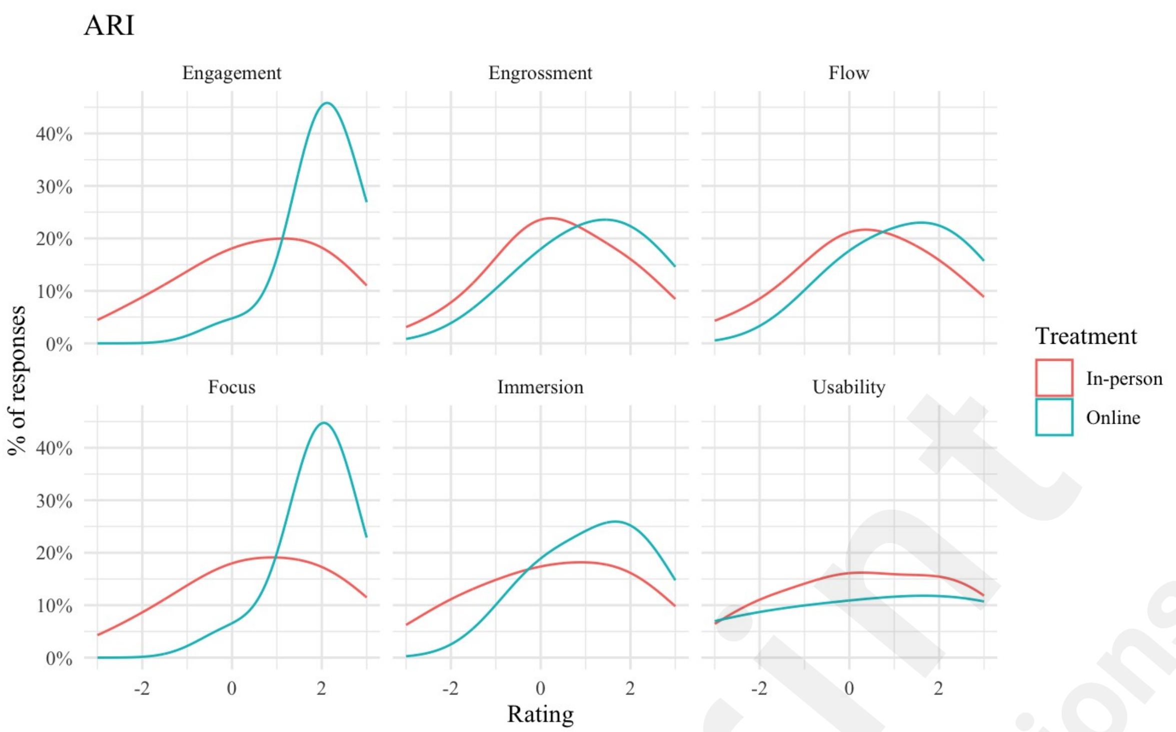


Figure 3: ARI results grouped by constructs from both the online participants and the student participants.

We also look at the user experience with the UEQ-S survey. As the scales are subjective in judgment, the developers of UEQ-S have provided a benchmark that compares the usability across the average of thousands of other usability studies. Both online and offline, the game was perceived positively for its hedonic quality but was lacking somewhat in the pragmatic quality compared to these other usability studies (Figure 4).

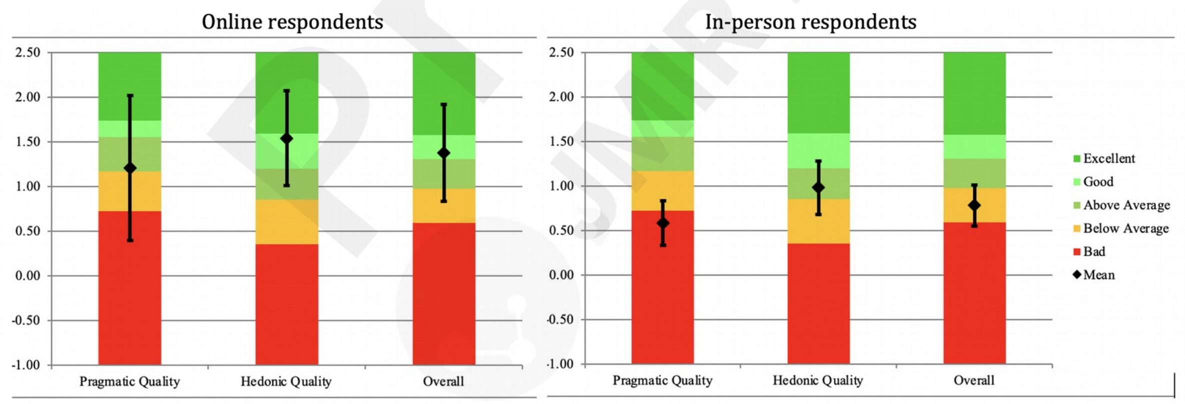


Figure 4: UEQ-S results from online respondents and in-person respondents. The online respondents have a more positive perception of AR osteoclasts placing its hedonic quality significantly above the mean of UEQ-S whereas in-person respondents only rate it slightly above average. The pragmatic quality is not significantly above average.

AR experience

The UEQ is not focusing on AR experiences, so we also gauge whether participants are able to access the AR environment continuously. We track how often they lose and find the page on which the AR scene is anchored (Figure 5). We opted for this feature so that people who focus on the page get the AR experience from the first moment. In the tutorial, the scene can be seen but is not the focus as the User Interface is explained. Only in the AR setup stage do we train participants how to focus on the page. After the training, we can see that for the rest of the time people effectively refocus on the page bringing us to net zero between pages lost and found for the AR tracking.

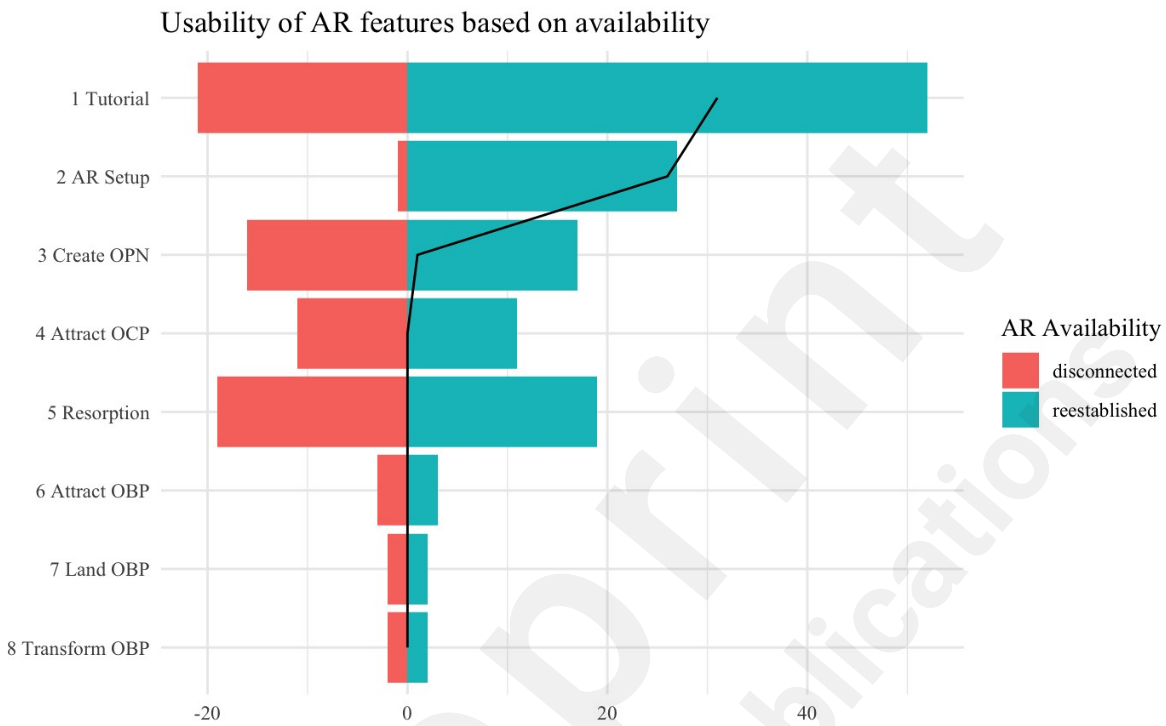


Figure 5: Usability of the AR features based on whether the AR was able to be displayed. Disconnected AR was caused when the camera could not see the book page. Connection was reestablished when the book page came into focus. During the Tutorial and the AR setup, most people acquire the AR display. During the game, the average is around 0 indicating that people were able to recover from losing the AR display.

Playtime

The completion time by stages of the game reveals three patterns of interest (Figure 6). The average play time starts off as a normal distribution but as the stages progress it transforms into a trimodal distribution. At the end of the game, we have three distinct peaks at t=3 minutes, t=8 minutes, and t=16 minutes. We will refer to these as short, average and extensive playtimes, respectively. These playtimes are indicative of playstyles [45] that either try to play the game as fast as possible, engage with the content as presented or try to explore every possible aspect of the game. The tail probabilities indicate that about half of the participants belong to the average playtime group (p=0.5) whereas both other groups have a tail probability of around 10% (p=0.1).

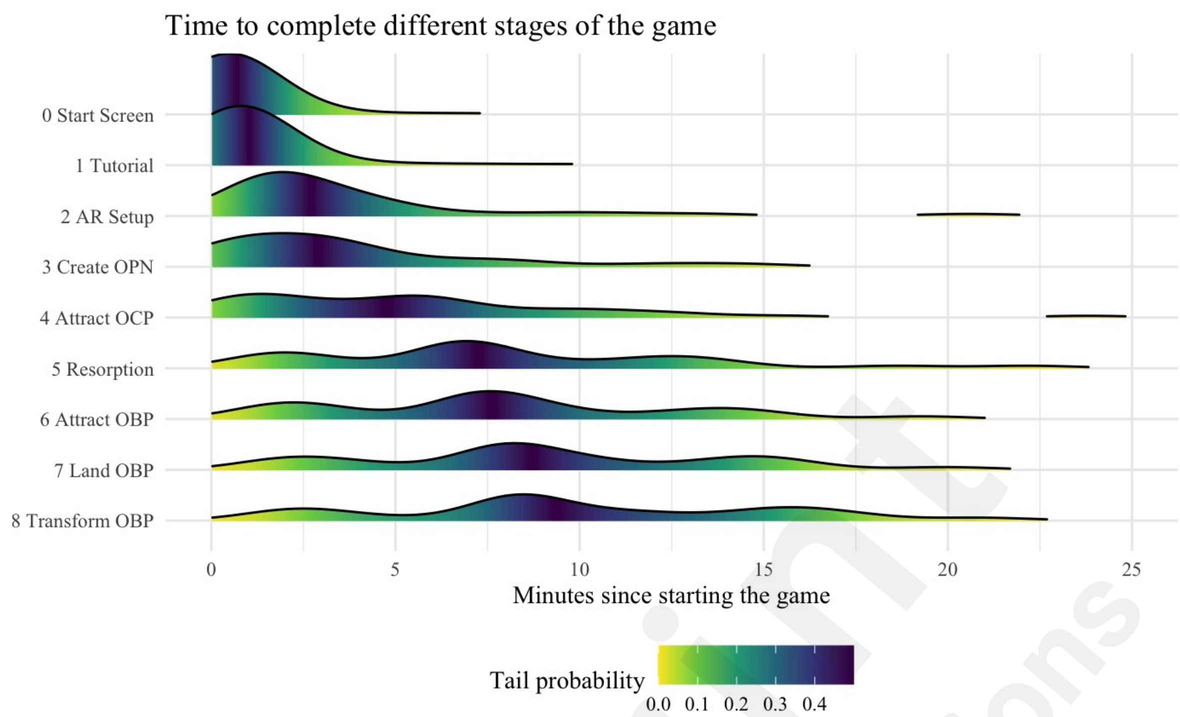


Figure 6: The stages of the game are shown from top to bottom. The time to complete is shown across all participants. The height of the ridges shows how many participants completed the stage at a specific minute. The colors are based on the empirical cumulative distribution function indicating that 95% of participants fall in the blue to green area and only 5% fall in the yellowish tones. We can see that the main cohort play the game in 8 minutes and that there are a few fast players completing it in 3 minutes and slow players completing the game in 16 minutes.

User actions

We show the user interactions by game stage to investigate how easily participants were able to engage with the content (Figure S4 Appendix A.3). We observe that most failures to interact happen in the first game stage. There are two potential causes, either people do not yet understand what to do or they are so curious about the novel environment that they randomly interact with anything. Furthermore, half the stages exhibit “miss” rates of nearly 50%. This could indicate that people had difficulties with the user interactions where they understood what to do but were not able to do it correctly. Finally, two stages have success rates above 80% indicating that participants understood the tasks and performed them correctly.

However, we would like to understand the failures and misses better. Therefore, we investigate the user interactions by playtime leaning on the trimodal distribution of timing (Figure 7). For short playtime, we find very low failure rates and distinctly higher success rates than for the overall group. For the average playtime, we find that most failures are concentrated in this group. Lastly, we find that the average player seems to get more used to the game mechanics in the next stage where they exhibit a high miss rate instead of a failure rate. Overall, many misses in the last stages are accrued by the extensive playtimes.

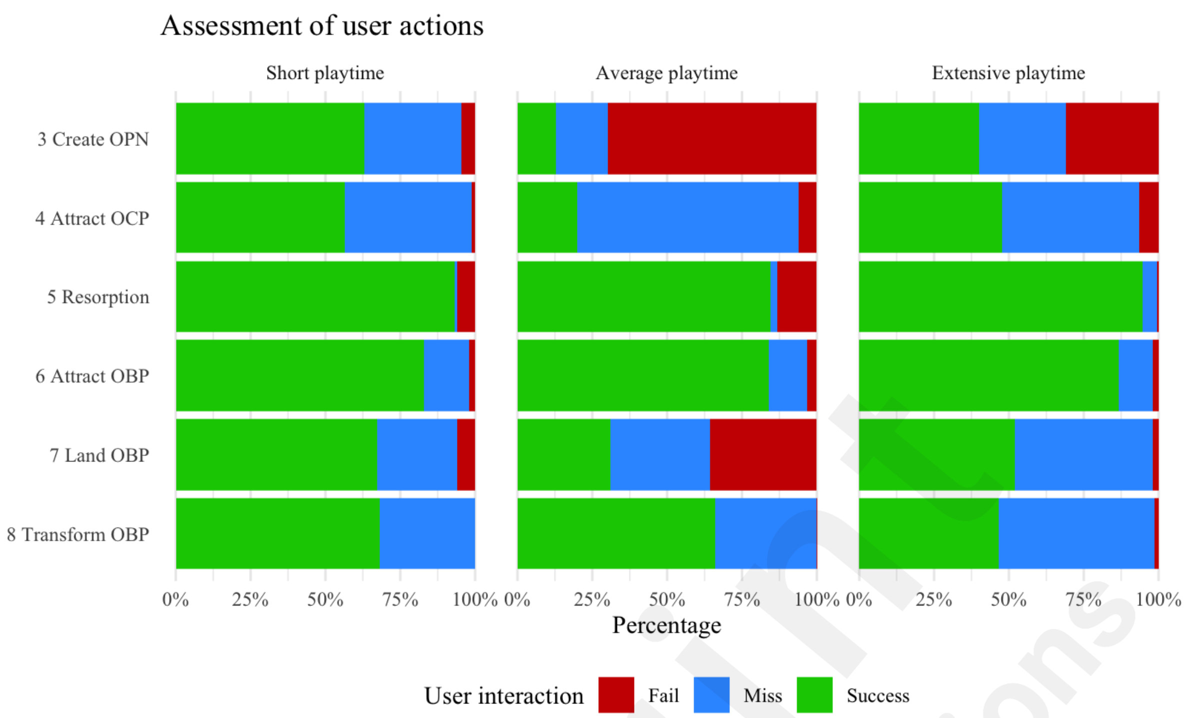


Figure 7: Short play times relate to higher success rates and extensive playtimes have higher miss rates. Most failures are occurring with average playtime, especially at the beginning of the game.

Student Learning Experiment

The student learning experiment gives us crucial insight into how much the students were able to learn from the serious game. We find support for H1a but at this point we find no indication for H1b. However, this study is possibly underpowered as some of the questions were not apt at discriminating between different levels of learning as they were too easy and might have induced a ceiling effect. For H2, we find evidence and for H3 we find from topic modeling that topics around cells are more prevalent in AR and topics about the process are more connected to the text. Lastly, we cannot fully assess H4 due to participant attrition of the long-term study. In summary, we find that overall students learn the key concept but that curious students benefit the most from AR experiences.

Student Statistics

39 Students (23 female, 14 male, 2 not registered because of attrition before the survey and preference not to answer) aged 27 (standard deviation = 3.6; measures in 5-year blocks). Students mostly heard of osteoclasts (36 students), and a few had worked with osteoclasts before (2 students). 6 students started studying dental medicine with the master's program (0-1 years in dental medicine), most students started with dental medicine in their bachelor's program (3 years of experience) and 6 students have extended experience (5 years of experience). Students also self-reported their CEI-II as 32.26 on average (standard deviation = 6.74).

Evaluation Outcomes

We evaluate the outcome of the multiple-choice questions for recall, transfer, and visual transfer tasks and the text responses of participants.

Multiple-choice questions

We report the participant responses as the proportions of correct answers for each question. For recall tasks (Appendix 1B, Figure S5), we report very high correct responses, indicating a ceiling effect. Whereas several questions have been answered correctly by 100% of the participants, there are minor differences between treatment groups. However, these appear to be random. For transfer tasks (Appendix 1B Figure S6), we observe that two questions appear simple with very high correct response rates similar to recall tasks. The other three questions have much correct response rates ranging from 50%-70%. Lastly, for visual transfer tasks (Appendix 1B Figure S7), we report the lowest correct response rates. A Welch t-test reveals that the AR treatment is not significantly different from the text treatment in terms of learning outcome (p=0.114). A comparison with the online sample exhibits the same characteristics (Appendix 1D). All transfer tasks have nominally higher scores for text treatments but not on a significant level.

In our regression analysis (Table 3), we report no significant results for recall and transfer learning. These models have adjusted R^2 of 0.03 and 0.12, respectively. Their f-statistics are not significant. The models for Visual Transfer and Total cumulative score have a significant f-statistic ($f_{visual\_transfer}$=2.78 and f_{total}=2.63, respectively). Only the visual transfer model reports significant effects for the AR treatment ($p_{visual\_transfer}$= 0.04). All other models cannot make a difference between the treatment conditions for the outcome variables. Across models, we also find that previous experience with osteoclasts is a strong predictor of the result (p_{total}=0.02, $p_{transfer}$=0.08, $p_{visual\_transfer}$=0.07). For the visual transfer task, we also find a significant effect of the CEI-II ($p_{visual\_transfer}$=0.02), the interaction effect ($p_{visual\_transfer}$=0.06) and the preference for text learning ($p_{visual\_transfer}$=0.02). Lastly, the cumulative score also finds a significant effect for the participants being male (p_{total}=0.06).

Table 3: Student outcomes for the three learning tasks and a cumulative score.
For each variable, the top row indicates the effect size and the bottom row the standard error.

	Dependent variable:			
	Total	Recall	Transfer	Visual Transfer
AR treatment	-0.211	0.079	-0.091	**-0.620****
	(0.175)	(0.208)	(0.362)	**(0.291)**
CEI-II	-0.001	0.005	0.005	**-0.014****
	(0.003)	(0.004)	(0.007)	**(0.006)**
ARI	0.004	0.005	0.009	-0.003
	(0.003)	(0.003)	(0.006)	(0.004)
Swiss grade	0.040	0.049	0.061	0.011
	(0.047)	(0.056)	(0.098)	(0.079)
Age	-0.006	0.004	-0.011	-0.012
	(0.005)	(0.006)	(0.011)	(0.009)
Male	**0.079***	0.084	0.074	0.080
	(0.041)	(0.049)	(0.085)	(0.069)

Experience	**0.194****	0.037	**0.310***	**0.236***
	(0.081)	(0.096)	**(0.167)**	**(0.134)**
Image preference	-0.011	0.002	-0.034	0.001
	(0.014)	(0.017)	(0.029)	(0.023)
Interaction preference	-0.004	-0.013	0.007	-0.006
	(0.015)	(0.018)	(0.031)	(0.025)
Text preference	0.016	-0.019	0.014	**0.054****
	(0.012)	(0.015)	(0.026)	**(0.021)**
Interaction of AR treatment and CEI-II	0.004	-0.005	-0.0002	**0.018***
	(0.005)	(0.006)	(0.011)	**(0.009)**
Constant	0.180	0.568	-0.180	0.160
	(0.440)	(0.522)	(0.909)	(0.732)
Observations	36	36	36	36
R^2	0.546	0.336	0.399	0.560
Adjusted R^2	0.338	0.032	0.124	0.359
Residual Std. Error (df = 25)	0.100	0.119	0.206	0.166
F Statistic (df = 10; 25)	**2.627****	1.104	1.452	**2.779****
Note:				*p<0.1; **p<0.05; ***p<0.01

We investigate the marginal effects of the model for visual transfer learning (Figure 8). In the range of the CEI-II index, we detect an inflection point at 36 points. For lower CEI-II scores, text-only learning yields the best results. However, for higher CEI-II scores, AR experiences help participants to score higher. The results are limited by the empirical distribution not having observations for CEI-II under 16 or over 42 making the behavior in the tails of the distribution speculative.

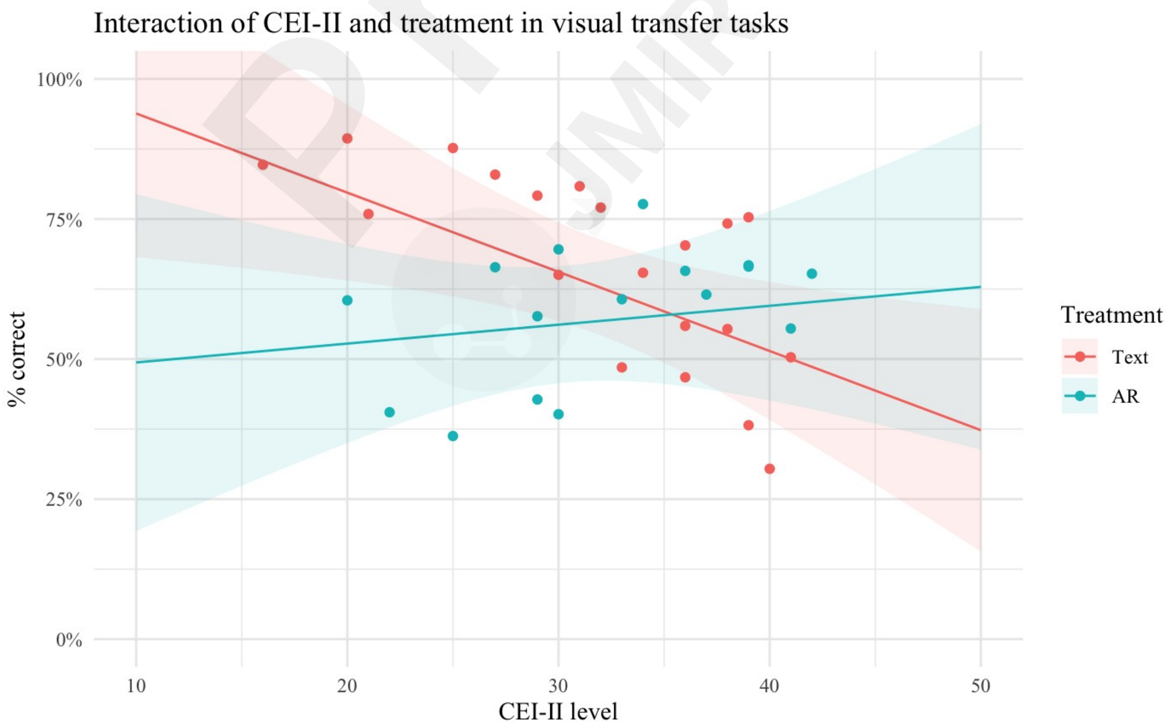


Figure 8: Our model reveals a strong inverse relationship between reading and curiosity in the

students. For students that are curious, this can be offset by experiencing the information in AR.

Due to attrition, we cannot report statistical results for the long- term study. However, those participants who did complete the follow-up, show remarkable performance when playing the AR game (Appendix 1B Figure S8). Some appear to have improved the retention over time. That being said, with such a low number of observations, it cannot be assessed further.

Text responses

We first illustrate the data with word clouds (Appendix 1B, Figure S9) to understand what participants responded. Generally, we find that the words correspond broadly to topics within the question. Describing the answers to the different questions, we see certain words can be assigned to groups. For a) formation of osteoclasts, the most common answers can be grouped to precursors (haemopoietic, progenitor, stem) and receptors (RANKL, RGD, bind). For b) function of osteoclasts and osteoblasts, the answers group according to the two cell types: osteoclasts (resorb, dissolve), osteoblast (form, build). For c) osteogenic differentiation separates into cell related words (bone cell, mesenchyme, osteoblast, osteoclast, stem) and signaling / receptor related words (wnt, bmp, RANKL, RANK).

Our topic models were labeled by three independent raters. Our intercoder-reliability reached Krippendorff's α=0.717. This indicates good intercoder-reliability. There is no full agreement but when there was no consensus, we used the most probable words from the topic contrasts (Appendix 1E) to determine a label in the following figures. We look at the prevalence of topics across all texts where 100% is all words written (Appendix 1B, Figure S10). We can see that the topic on osteogenic differentiation and RANKL signaling are most talked about with each having about 15% or more of the texts. The topics on differentiated cells, remodeling site and stem cell each garnered more than 10% of the total text. The topics of bone resorption mechanism, RANK signaling, bone remodeling and multinucleated cells each gathered more than 5% of the total text. The smallest topic is the "Don't know" topic that gathered uncertain responses by students.

We use the properties of structural topic modeling to assess whether our treatment is driving students towards different topics (Figure 9). We find that in the AR treatment students write more about the stem cells and differentiated cells whereas in the text treatment the focus is on the remodeling site and the process of bone remodeling. People in the text-only treatment more often respond with "Don't know". All other topics do not exhibit significant differences between treatments.

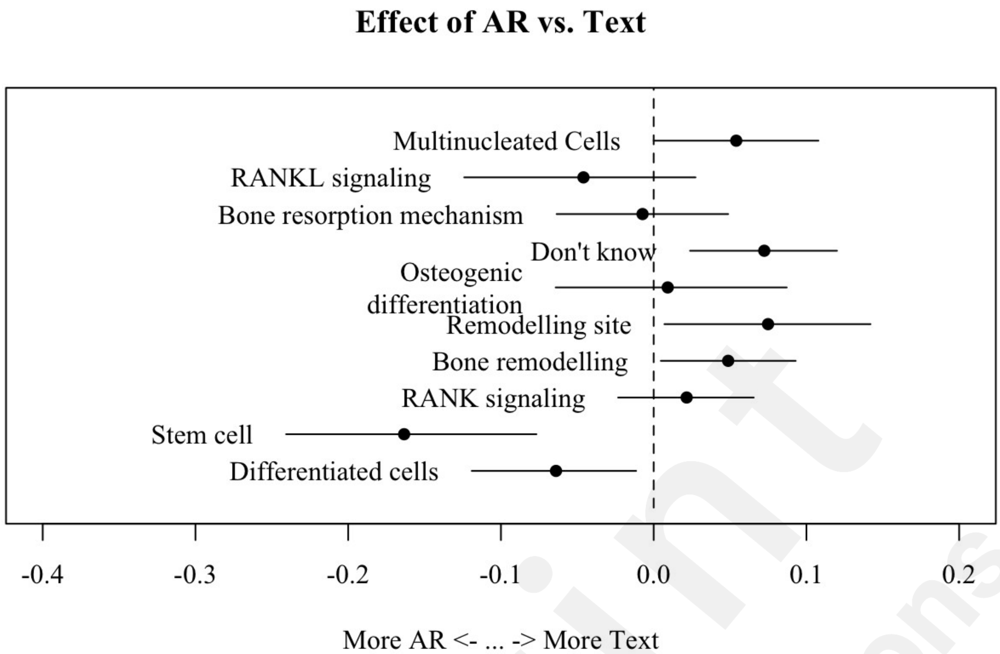


Figure 9: Topic prevalence by the treatments. Negative values indicate a leaning towards the AR treatment, positive values indicate a leaning towards the text-only treatment. If the dashed line is touched, there is no significant leaning towards either side detected.

Topics may also be driven by the three questions we asked (Appendix 1C.3 Figure S11). To gauge whether a topic belongs to a specific question, we need to check the topic in all three paired comparisons between questions. For instance, the topic on multinucleated cells is associated less often with the question on osteoclast formation (Q1) but is strongly associated with both coordinated function of osteoclasts and osteoblast in the canopy (Q2) and osteogenic differentiation (Q3). That means participants are statistically significantly more likely to talk about multinucleated cells in Q2 and Q3.

In the formation question (Q1), participants are less likely to talk about RANK-RANKL signaling and stem cells and multinucleated cells and more likely to talk about bone remodeling, bone resorption, the remodeling site, differentiated cells, stem cells, or osteogenic differentiation. In the functions question (Q2), participants are less likely to talk about bone remodeling, bone resorption mechanism, the remodeling site, or differentiated cells and more likely to talk about osteogenic differentiation, RANK-RANKL signaling and stem cells and multinucleated cells. In the differentiation question (Q3), participants are less likely to talk about bone resorption mechanism, stem cells, RANK signaling, or osteogenic differentiation and more likely to talk about bone resorption, the remodeling site, differentiated cells, multinucleated cells and RANKL signaling.

Subjective Learning

For 8 questions, the IMMS showed significant differences between study groups (Appendix F Figure S29). Student in the AR group experienced the content as more useful to their study (content useful: *"The content of this lesson will be useful to me."; p<0.001*), more organized (content structure: *"The good organization of the content helped me be confident that I would learn this material.";*

p=0.049), and more worthy of study (appears important: *"The content and style of writing in this lesson convey the impression that its content is worth knowing."; p=0.032*). Students also perceived the presentation as more appealing (eye catching: *"These materials are eye-catching."; p=0.047*) and easier to learn (impression easy: *"When I first looked at this lesson, I had the impression that it would be easy for me."; p=0.044*). They were also more interested to learn more about the topic (piqued curiosity: *"This lesson has things that stimulated my curiosity."; P=0.038*) and less turned off by the presentation (too dry: "The pages of this lesson look dry and unappealing."*; p=0.037*). Students also felt more accomplished completing the AR game (feel successful: *"It felt good to successfully complete this lesson."; p=0.008*). In the ARCS model (Figure 10), we find significant effects for Confidence (p=0.023), Relevance (p<0.001), and Satisfaction (p<0.001). Participants do not perform better, but they are overall more motivated. In particular, they feel more confident, they feel like this is a relevant topic to learn, and they are satisfied with their progress. Interestingly, we did not find significant differences in Attention (p=0.084). This is surprising as previous interactive learning tools are known to increase attention [46,47]. However, this could be explained by our small sample size.

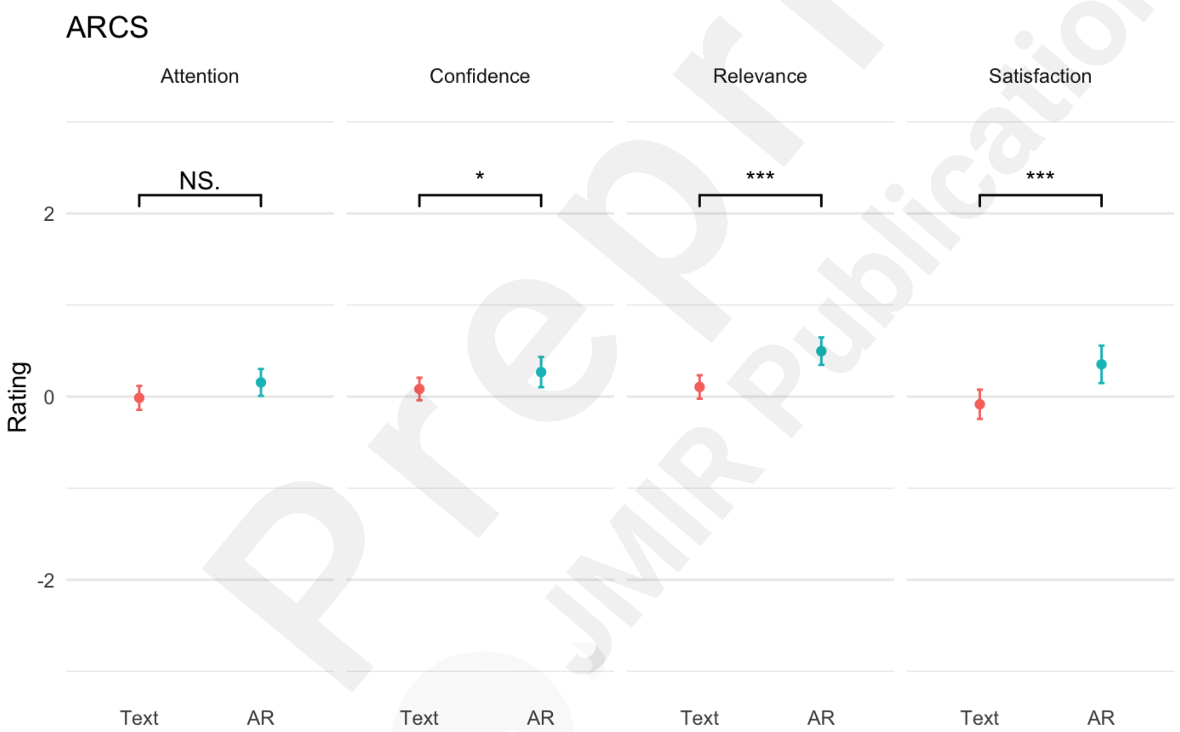


Figure 10: Evaluating the IMMS responses according to the ARCS construct we find that Attention (p=0.084) does not differ between learning scenarios, but Confidence (p=0.023) is increased, the material is perceived with more Relevance (p<0.001), and student experience more Satisfaction (p<0.001) learning.

Discussion

In this study we evaluated the application of an AR medical teaching game in a classroom setting. We study learning outcomes with a two-treatments design. A group of students receives the AR serious game in combination with a written text and is compared to a second group, which only receives the text. The aim was to analyze whether an AR game can positively influence the engagement of students in a new topic and to give them an immersive experience. Further we investigated whether long-term knowledge retention can be improved via this AR tool. For this purpose, the game was accessible worldwide. This allowed a worldwide community to try and play the game, giving us broad feedback. We find encouraging signs for the positive effect of immersive media in education with the limitations that curious students benefit most from interactive learning media.

Principal Results

We find that presenting advanced scientific content, in our case the process of bone remodeling, in Augmented Reality (AR) helps students to engage with material they are learning and develop a deeper appreciation of the topic and perceive the topic as overall more important. Students in the AR experience also achieve comparable learning outcomes (H1a). However, students that rank high in curiosity even appear to perform better on visual transfer tasks with an AR experience than from a textbook alone (H2). We also find that the principal mechanisms in the serious game (the cells) drive the text responses of participants and shape how participants learn and memorize content (H3).

To disentangle the learning from the playing experience, we investigated both the user experience of the app and the subjective learning experience of the students. We find that in-class participants rated the user experience as less new than online participants who were overall more positive towards the application. We explain this difference by the fact that the students took part in the experiment by default whereas online participants approached the game on their own volition. Generally, the hedonic quality of the game was rated higher compared to the pragmatic quality. We interpret this to indicate that the game was visually pleasing and a positive experience. But sometimes the game mechanics may have been too loosely related to the scientific process that they represent.

Online participants have allowed us to gain additional insights into users by increasing the number of participants from 19 to a total of 376 before attrition. The game attracted the most activity in Germany, India, the U.S., Chile, Switzerland and Saudi Arabia. We attribute the localities to the distribution of the accompanying book and the prevalence of the book as a tool in education and practice.

We aimed for users to take about 10 minutes to complete the AR experience. Most participants played the game for approximately 8 minutes with a trimodal distribution of playtime with minor peaks at 3 minutes and 16 minutes indicating different playstyles for so-called speed runners (i.e. completing the game as fast as possible) to completionists (i.e. exploring every possible aspect of a game) [48]. Keeping playstyles in mind may be important to improve participant engagement by catering to their different interactions with the game. We find that people have no problems with the AR setup of the game. However, our analysis of the user actions also indicates that average playtime was associated with most difficulties to operate the game. One reason for this could be different prior knowledge of AR and tech-savviness. Experienced players tend to have less trouble operating a game [28] and can therefore explore more efficiently or more broadly.

We tested the learning outcomes for tasks ranging from recall to transfer learning and visual transfer learning. We find no significant differences between the AR experience and text-only treatment. In effect, the AR experience provides an equivalent opportunity to learn overall. However, we found two further advantages of the AR experience. First, students who are curious according to the CEI-II seem to benefit from the AR experience when being confronted with visual transfer tasks. We find a cutoff at CEI-II= 35 beyond which students start to perform better on visual transfer tasks after playing the serious game. As dental medicine in its applied form requires interpretation of visual tasks, AR representations harbor a great opportunity. Second, the subjective learning experience indicates that students who experienced the AR scene, attribute higher importance to the learned concepts and find them relevant to learn. They also form a deeper personal connection with the topic as they attribute it to higher personal utility. The playful experience also seems to bring students to perceive the material as being easier to learn even if they did not improve in grading. This may allow students to experience less pressure while learning and feel more satisfied with their performance.

We also investigate the student's text responses to get a better understanding of the actual learning achievements. Interestingly, the students in the serious game treatment group mention the cells more compared to the text responses that tend to focus a bit more on processes and location. The principal features of the game in terms of interactivity and game mechanics seem to direct the students' wording in answering questions. The choice of how to represent the scientific process as a game is crucially forming the learning experience of the students. While the long-term study has limited generalizability, we observe that more general topics increase in topics whereas very specific topics (RANK-RANKL signaling) decrease in prevalence. Most topics maintained their prevalence, and this may point to some stable core of learned concepts that loses a bit of detail over time.

Limitations

We present four major limitations of this study that may be used as directives for future research and guides for how to improve future research on serious gaming in AR. Serious gaming in AR for education has to pay special attention to the (student) player characteristics, the transformation of scientific knowledge into game mechanisms, the effectiveness of long-term measurements, and the suitability of the tests to assess the game.

First on player characteristics. Games are not played in a vacuum by learners but by people with diverse backgrounds and gaming skills. Casual gamers [49] are often less familiar with AR experiences compared to other participants that prefer speedruns or completionist approaches [48], see Figure 18. Often, content caters to advanced players and may be lost on casual gamers. To ensure that a serious game can achieve its goal of imparting knowledge, we need to account for play styles in the game design. We observe that most participants were casual gamers indicating that performance issues in the average playtime group we found is more an issue of the user interface and game dynamics than the actual learning content [28].

By mechanizing a scientific process into a game, we need to make choices about what to represent and what to omit. We also decide which aspect of the process gets highlighted (visually, interactively, and mechanically) and which aspects are only indirectly present. We find that participants in AR talk more about the cells whereas participants with text-only focus on the location and process. The focus of the topics indicates that the AR leads students to engage more closely with the main mechanisms of the game, i.e. the cells, but also leads to a relative distancing of other aspects such as the overall process. With the strong visual guidance of the game, it is important to ensure that the game communicates not only the knowledge by solely presenting it but also by highlighting as part of the gaming experience. Another aspect of the text responses is the prevalence of the don’t know topic in the text-only treatment. We cannot conclude for certain whether students remember less in the text-only treatment or whether the students in the AR treatment decide to write something because of the vivid impressions of the game even when they do not know what they should write. To ascertain whether games can produce a more positive relation with the learned content it may be necessary to design studies that focus on this particular aspect.

The small number of participants and attrition for the longitudinal study limit our ability to draw conclusions in the long-term. However, some intriguing results could be observed that support further investigations. Selected students improved their knowledge retention over time when being exposed to the AR treatment. Whether these should be attributed to individual differences, or the treatment cannot be assessed with the collected data. The interaction of curiosity and AR treatment looks intriguing as well, but the small number of participants requires further investigation to detect solid proof for a positive relation between curiosity and gamified educational material.

Another limitation is the choice of tests. The questions for the three testing areas of recall, transfer and visual transfer need to be selected carefully to both match the content and be strong enough to elicit different responses across treatments. Ultimately, the questions may not have been difficult enough (ceiling effect) to produce a strong difference between participants at our expected effect sizes and power analysis. Most participants scored 75% correct or more for recall and transfer and the strongest predictor was previous knowledge of osteoclasts. The impact of AR on the ability to learn may not have been elicited beyond the visual transfer task.

Comparison with Prior Work

We identified 21 investigations of serious gaming for learning in a medical context that informed our research and to whose results we add new insights with our findings. In general, there is still a mixed view on the efficacy of (AR) serious gaming as several studies reported null findings whereas others reported benefits of gamifying the learning process. There are also positive reports on process learning with inverted classroom approach [50] to prepare students for their tasks as well as a generally higher acceptance of the content by the student and more positive experience of learning itself [51,52]. Another recent study from 2023 [53] showed that game-based learning is gradually increasing in education. The authors report that these teaching formats are well-received by learners and can create an immersive experience for students. We similarly find positive effects on intrinsic benefits, and we may offer an explanation as to why some studies have null-finding and others have not.

Null-findings were reported in a study in 2017 [54], a meta-review of 13 studies [55] and another meta-review of 27 studies [56] and a systematic review [57] of seven studies in physiotherapy with no significant benefits in the use of Virtual Reality (VR) or AR for learning. However, they equally suggested that VR and AR can promote intrinsic benefits such as increasing learner immersion and engagement. We can replicate positive effects on intrinsic benefits, but we also identified a potential moderator effect. The curiosity to engage the gaming content crucially drives visual understanding and only if tests capture these criteria can benefits be detected.

Studies that focus on the spatial abilities of medical students often find benefits from the embodied experience in VR/AR [58–62]. In medical education, a study performed in 2021 [63], comparing the teaching efficiency between virtual reality and traditional education. The results show that the pass rate of students using VR was significantly higher than those receiving traditional medical education. Apparently, the students developed a better understanding of different processes and procedures. In a study Weeks at al. [64] investigated whether AR could improve short-term anatomic knowledge in the head and neck anatomy compared to traditional 2D screen-based learning. The results indicate that the AR group performed significantly better. A further study [56] was able to show that VR/AR teaching had not only beneficial effects on medical education, but also showed that this type of teaching had a positive effect on enthusiasm and enjoyment of the participants. In our study we could show that visual transfer tasks are also positively affected by AR-based learning. This may be due to also eliciting spatial understanding in the medical context.

Conclusions

Immersive serious gaming in Augmented Reality is an important new tool for education in medical sciences and beyond. Scientific content is often hard to communicate and may scare off especially new students. Immersive media offer an engaging opportunity for students to familiarize themselves with complex learning material. We solidify previous findings that students build up a stronger connection with the content and perceive it more positively. For easy tasks, in line with the literature, we do not find learning benefits when using AR to create immersive experiences but for complex visual transfer tasks, we find that especially curious students benefit from AR-based learning. We contribute to the literature by identifying a potential pathway via curiosity when immersive experiences are especially fruitful for the student learning. Immersive serious gaming provides students with new perspectives on the content and for a subgroup of curious students can even influence their learning outcomes.

Acknowledgements

The game development was funded by Quintessential Publishing. 3D models were provided by interActive Systems. J.G. was supported by an ETH Zurich Doc.Mobility Fellowship. J.C. was supported by the Future Learning Initiative of ETH Zurich. The students of the 2022 cohort at UZH dental medicine are thanked for their participation in the study. We thank everybody that downloaded AR Osteoclasts and played the game, your participation has been instrumental in this research.

Conflicts of Interest

Reinhard Gruber and Bernd Stadlinger are together with Hendrik Terheyden editors of the book “Visual Biology in Oral Medicine”.

Abbreviations

2D: 2-dimensional
3D: 3-dimensional
AR: Augmented Reality
ARI: Augmented Reality Immersion

BMP: Bone Morphogenetic Proteins
CEI-II: Curiosity and Exploration Inventory-II
IMMS: Instructional Materials Motivation Survey
JMIR: Journal of Medical Internet Research
OB: Osteoblast
OBP: Osteoblast Progenitor / Precursor
OC: Osteoclast
OCP: Osteoclast Progenitor / Precursor
RANK: Receptor Activator of NF-κB
RANKL: Receptor Activator of NF-κB Ligand
RCT: randomized controlled trial
RGD: Arg-Gly-Asp (Arginylglycylaspartic acid)
VEGF: Vascular Endothelial Growth Factor
VR: Virtual Reality
WNT: Wingless und Int-1 (Integrated)
UEQ: User Experience Questionnaire
UEQ-S: User Experience Questionnaire - Short

References

1. McConnell DA, Chapman L, Czajka CD, Jones JP, Ryker KD, Wiggen J. Instructional utility and learning efficacy of common active learning strategies. J Geosci Educ Informa UK Limited; 2017 Nov;65(4):604–625. doi: 10.5408/17-249.1
2. Yu J, Denham AR, Searight E. A systematic review of augmented reality game-based Learning in STEM education. Educ Technol Res Dev Springer; 2022 Aug 1;70(4):1169–1194. doi: 10.1007/s11423-022-10122-y
3. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. Telemanipulator and Telepresence Technologies SPIE; 1995. p. 282–292. doi: 10.1117/12.197321
4. Scaife M, Rogers Y. External cognition: how do graphical representations work? Int J Hum Comput Stud Elsevier; 1996 Aug 1;45(2):185–213. doi: 10.1006/ijhc.1996.0048
5. Gruber R, Stadlinger B. Cell-Atlas - Visual Biology in Oral Medicine. Quintessenz Verlag; 2022. Available from: https://play.google.com/store/books/details?id=gPBmEAAAQBAJ ISBN:9783868676181
6. Insua A, Monje A, Wang H-L, Miron RJ. Basis of bone metabolism around dental implants during osseointegration and peri-implant bone loss. J Biomed Mater Res A 2017 Jul;105(7):2075–2089. PMID:28281321
7. Eng-Tat A, Liu P, R J, M C, R L S, Lian WQD, Huppertz B, Khamuani M, Minarcik JR. The challenges of histology education. Clin Anat 2023 Mar;36(2):308–314. PMID:36494890
8. García M, Victory N, Navarro-Sempere A, Segovia Y. Students' Views on Difficulties in Learning Histology. Anat Sci Educ 2019 Sep;12(5):541–549. PMID:30378295
9. Taylor DCM, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE Guide No. 83. Med Teach 2013 Nov;35(11):e1561–72.

PMID:24004029
10. Graffam B. Active learning in medical education: strategies for beginning implementation. Med Teach 2007 Feb;29(1):38–42. PMID:17538832
11. Sait MS, Siddiqui Z, Ashraf Y. Advances in medical education and practice: student perceptions of the flipped classroom. Adv Med Educ Pract. 2017. p. 317–320. PMID:28496377
12. Kapur M, Bielaczyc K. Designing for Productive Failure. Journal of the Learning Sciences Routledge; 2012 Jan 1;21(1):45–83. doi: 10.1080/10508406.2011.591717
13. Felszeghy S, Pasonen-Seppänen S, Koskela A, Nieminen P, Härkönen K, Paldanius KMA, Gabbouj S, Ketola K, Hiltunen M, Lundin M, Haapaniemi T, Sointu E, Bauman EB, Gilbert GE, Morton D, Mahonen A. Using online game-based platforms to improve student performance and engagement in histology teaching. BMC Med Educ 2019 Jul 22;19(1):273. PMID:31331319
14. Dhar P, Rocks T, Samarasinghe RM, Stephenson G, Smith C. Augmented reality in medical education: students' experiences and learning outcomes. Med Educ Online 2021 Dec;26(1):1953953. PMID:34259122
15. Kwon H-B, Park Y-S, Han J-S. Augmented reality in dentistry: a current perspective. Acta Odontol Scand 2018 Oct;76(7):497–503. PMID:29465283
16. De Micheli AJ, Valentin T, Grillo F, Kapur M, Schuerle S. Mixed Reality for an Enhanced Laboratory Course on Microfluidics. J Chem Educ American Chemical Society; 2022 Mar 8;99(3):1272–1279. doi: 10.1021/acs.jchemed.1c00979
17. Sweller J, Ayres P, Kalyuga S. The Split-Attention Effect. In: Sweller J, Ayres P, Kalyuga S, editors. Cognitive Load Theory New York, NY: Springer New York; 2011. p. 111–128. doi: 10.1007/978-1-4419-8126-4_9
18. Hornecker E. The to-and-fro of sense making. ACM Trans Comput Hum Interact Association for Computing Machinery (ACM); 2016 May 28;23(2):1–48. doi: 10.1145/2882785
19. Rothbaum BO, Rizzo AS, Difede J. Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Ann N Y Acad Sci 2010 Oct;1208:126–132. PMID:20955334
20. Molnar A. Content type and perceived multimedia quality in mobile learning. Multimed Tools Appl 2017 Oct 1;76(20):21613–21627. doi: 10.1007/s11042-016-4062-2
21. Zünd F, Ryffel M, Magnenat S, Marra A, Nitti M, Kapadia M, Noris G, Mitchell K, Gross M, Sumner RW. Augmented creativity: bridging the real and virtual worlds to enhance creative play. SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications New York, NY, USA: Association for Computing Machinery; 2015. p. 1–7. doi: 10.1145/2818427.2818460
22. Ryffel M, Zünd F, Aksoy Y, Marra A, Nitti M, Aydın TO, Sumner B. AR Museum: A mobile augmented reality application for interactive painting recoloring. ACM Trans Graph disneyresearch.s3.amazonaws.com; 2017;36(2):19. Available from: http://disneyresearch.s3.amazonaws.com/wp-content/uploads/20170814151522/AR-Museum-A-Mobile-Augmented-Reality-Application-for-Interactive-Painting-Recoloring-Paper.pdf
23. Dzyuba N, Jandu J, Yates J, Kushnerev E. Virtual and augmented reality in dental education: The good, the bad and the better. Eur J Dent Educ 2022 Nov 6; PMID:36336847
24. Yoo S, Heo S, Song S, Park A, Cho H, Kim Y, Cha WC, Kim K, Son MH. Adoption of Augmented Reality in Educational Programs for Nurses in Intensive Care Units of Tertiary Academic Hospitals: Mixed Methods Study. JMIR Serious Games JMIR Serious Games; 2024 May 23;12(1):e54188. doi: 10.2196/54188

25. Gasteiger N, van der Veer SN, Wilson P, Dowding D. How, for Whom, and in Which Contexts or Conditions Augmented and Virtual Reality Training Works in Upskilling Health Care Workers: Realist Synthesis. JMIR Serious Games JMIR Serious Games; 2022 Feb 14;10(1):e31644. doi: 10.2196/31644
26. Gruber R. Osteoimmunology: Inflammatory osteolysis and regeneration of the alveolar bone. J Clin Periodontol 2019 Jun;46 Suppl 21:52–69. PMID:30623453
27. Terheyden H, Lang NP, Bierbaum S, Stadlinger B. Osseointegration--communication of cells. Clin Oral Implants Res Wiley; 2012 Oct;23(10):1127–1135. PMID:22092345
28. Gruber R, Stadlinger B, Terheyden H. Cell-to-cell communication in guided bone regeneration: molecular and cellular mechanisms. Clin Oral Implants Res 2017 Sep;28(9):1139–1146. PMID:27550738
29. Stadlinger B, Jepsen S, Chapple I, Sanz M, Terheyden H. Technology-enhanced learning: a role for video animation. Br Dent J 2021 Jan;230(2):93–96. PMID:33483667
30. Grübel J, Thrash T, Hölscher C, Schinazi VR. Evaluation of a conceptual framework for predicting navigation performance in virtual reality. PLoS One journals.plos.org; 2017 Sep 15;12(9):e0184682. PMID:28915266
31. Roberts ME, Stewart BM, Tingley D. stm: An R Package for Structural Topic Models. J Stat Softw jstatsoft.org; 2019 Oct 31;91:1–40. doi: 10.18637/jss.v091.i02
32. Laugwitz B, Held T, Schrepp M. Construction and Evaluation of a User Experience Questionnaire. HCI and Usability for Education and Work Springer Berlin Heidelberg; 2008. p. 63–76. doi: 10.1007/978-3-540-89350-9_6
33. Schrepp M, Hinderks A, Thomaschewski J. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). Int J Interact Multimed Artif Intell IMAI Software - International Journal of Interactive Multimedia and Artificial Intelligence; 2017;4(6):103. doi: 10.9781/ijimai.2017.09.001
34. Schrepp M. A Comparison of SUS, UMUX-LITE, and UEQ-S. uxpajournal.org; 2023. Available from: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUX_Schrepp_Feb2023_updated.pdf [accessed May 13, 2024]
35. Georgiou Y, Kyza EA. The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. Int J Hum Comput Stud Elsevier; 2017 Feb 1;98:24–37. doi: 10.1016/j.ijhcs.2016.09.014
36. Kashdan TB, Gallagher MW, Silvia PJ, Winterstein BP, Breen WE, Terhar D, Steger MF. The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics. J Res Pers 2009 Dec 1;43(6):987–998. PMID:20160913
37. Perkins DN, Salomon G, Others. Transfer of learning. International encyclopedia of education researchgate.net; 1992;2:6452–6457. Available from: https://www.researchgate.net/profile/Gavriel-Salomon/publication/2402396_Transfer_Of_Learning/links/553506410cf2df9ea6a42d77/Transfer-Of-Learning.pdf
38. Rau MA. Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. Educ Psychol Rev Springer; 2017 Dec 1;29(4):717–761. doi: 10.1007/s10648-016-9365-3
39. Loorbach N, Peters O, Karreman J, Steehouder M. Validation of the Instructional Materials

Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. Br J Educ Technol Wiley; 2015 Jan;46(1):204–218. doi: 10.1111/bjet.12138

40. Keller JM. Development of two measures of learner motivation. Unpublished Manuscript in progress Florida State University 2006;
41. Hainmueller J, Mummolo J, Xu Y. How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. Polit Anal Cambridge University Press; 2019 Apr;27(2):163–192. doi: 10.1017/pan.2018.46
42. Arel-Bundock V. Marginal Effects. Marginal Means, Predictions, and Contrasts Available online: https://CRAN R-project org/package= marginaleffects (accessed on 10 April 2023) 2021;
43. Seery K, Barreda A, Hein S, Hiller J. Retention strategies for online students: A systematic literature review. Journal of Global Education and Research University of South Florida Libraries; 2021 Jun;5(1):72–84. doi: 10.5038/2577-509x.5.1.1105
44. Krippendorff K. Content Analysis: An Introduction to Its Methodology. SAGE Publications; 2018. Available from: https://play.google.com/store/books/details?id=nE1aDwAAQBAJ ISBN:9781506395678
45. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? BMC Med Res Methodol 2016 Aug 5;16(1):93. doi: 10.1186/s12874-016-0200-9
46. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin [International Biometric Society, Wiley]; 1945;1(6):80–83. doi: 10.2307/3001968
47. Jaćević M. How the Players Get Their Spots: A Study of Playstyle Emergence in Digital Games. 2021 IEEE Conference on Games (CoG) IEEE; 2021. p. 1–8. doi: 10.1109/CoG52621.2021.9619067
48. Wilujeng I. Increasing the Attention, Relevance, Confidence and Satisfaction (ARCS) of Students through Interactive Science Learning Multimedia. Research in Learning Technology 2021;29. doi: 10.25304/rlt.v29.2383
49. Chatain J, Varga R, Fayolle V, Kapur M, Sumner RW. Grounding Graph Theory in Embodied Concreteness with Virtual Reality. Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction New York, NY, USA: Association for Computing Machinery; 2023. p. 1–13. doi: 10.1145/3569009.3572733
50. O'Brien SP. "Select the Type of Experience You Would Like to Have": Exploring Player Roles and Role Affordance in Video Games. paj: The Journal of the Initiative for Digital Humanities, Media, and Culture paj-ojs-tamu.tdl.org; 2012;3. Available from: https://paj-ojs-tamu.tdl.org/paj/article/view/38
51. Kuittinen J, Kultima A, Niemelä J, Paavilainen J. Casual games discussion. Proceedings of the 2007 conference on Future Play New York, NY, USA: Association for Computing Machinery; 2007. p. 105–112. doi: 10.1145/1328202.1328221
52. Sulong A, Ibrahim AB, Abas A, Bakar AZA. Incorpating gamification in a flipped classroom approach: A review of literature. JPB ojs.upsi.edu.my; 2021 Apr 28;14:22–32. doi: 10.37134/bitara.vol14.sp.3.2021
53. Moreno-Ger P, Torrente J, Bustamante J, Fernández-Galaz C, Fernández-Manjón B, Comas-Rengifo MD. Application of a low-cost web-based simulation to improve students' practical skills in medical education. Int J Med Inform 2010 Jun;79(6):459–467. PMID:20347383

54. Huber J, Witti M, Schunk M, Fischer MR, Tolks D. The use of the online Inverted Classroom Model for digital teaching with gamification in medical studies. GMS J Med Educ ncbi.nlm.nih.gov; 2021 Jan 28;38(1):Doc3. PMID:33659608
55. Xu M, Luo Y, Zhang Y, Xia R, Qian H, Zou X. Game-based learning in medical education. Front Public Health 2023 Mar 3;11:1113682. PMID:36935696
56. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549–559. PMID:28419750
57. Baashar Y, Alkawsi G, Ahmad WNW, Alhussian H, Alwadain A, Capretz LF, Babiker A, Alghail A. Effectiveness of Using Augmented Reality for Training in the Medical Professions: Meta-analysis. JMIR Serious Games JMIR Serious Games; 2022 Jul 5;10(3):e32715. doi: 10.2196/32715
58. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. JMIR Serious Games 2021 Jul 8;9(3):e29080. PMID:34255668
59. Lucena-Anton D, Fernandez-Lopez JC, Pacheco-Serrano AI, Garcia-Munoz C, Moral-Munoz JA. Virtual and Augmented Reality versus Traditional Methods for Teaching Physiotherapy: A Systematic Review. Eur J Investig Health Psychol Educ 2022 Dec 2;12(12):1780–1792. PMID:36547026
60. Jung E-Y, Park DK, Lee YH, Jo HS, Lim YS, Park RW. Evaluation of practical exercises using an intravenous simulator incorporating virtual reality and haptics device technologies. Nurse Educ Today 2012 May;32(4):458–463. PMID:21664014
61. Hashimoto DA, Petrusa E, Phitayakorn R, Valle C, Casey B, Gee D. A proficiency-based virtual reality endoscopy curriculum improves performance on the fundamentals of endoscopic surgery examination. Surg Endosc 2018 Mar;32(3):1397–1404. PMID:28812161
62. Yoganathan S, Finch DA, Parkin E, Pollard J. 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. Int J Surg 2018 Jun 1;54:24–27. doi: 10.1016/j.ijsu.2018.04.002
63. Maytin M, Daily TP, Carillo RG. Virtual reality lead extraction as a method for training new physicians: a pilot study. Pacing Clin Electrophysiol 2015 Mar;38(3):319–325. PMID:25494952
64. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg Elsevier; 2007 Aug;194(2):205–211. PMID:17618805
65. Zhao G, Fan M, Yuan Y, Zhao F, Huang H. The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis. Ann Transl Med 2021 Feb;9(3):252. PMID:33708879
66. Weeks JK, Pakpoor J, Park BJ, Robinson NJ, Rubinstein NA, Prouty SM, Nachiappan AC. Harnessing Augmented Reality and CT to Teach First-Year Medical Students Head and Neck Anatomy. Acad Radiol 2021 Jun;28(6):871–876. PMID:32828663
67. Chen J. Flow in games (and everything else). Commun ACM New York, NY, USA: Association for Computing Machinery; 2007 Apr 1;50(4):31–34. doi: 10.1145/1232743.1232769
68. Grübel J, Thrash T, Aguilar L, Gath-Morad M, Chatain J, Sumner RW, Hölscher C, Schinazi VR. The hitchhiker’s guide to Fused Twins: A review of access to Digital Twins in situ in Smart

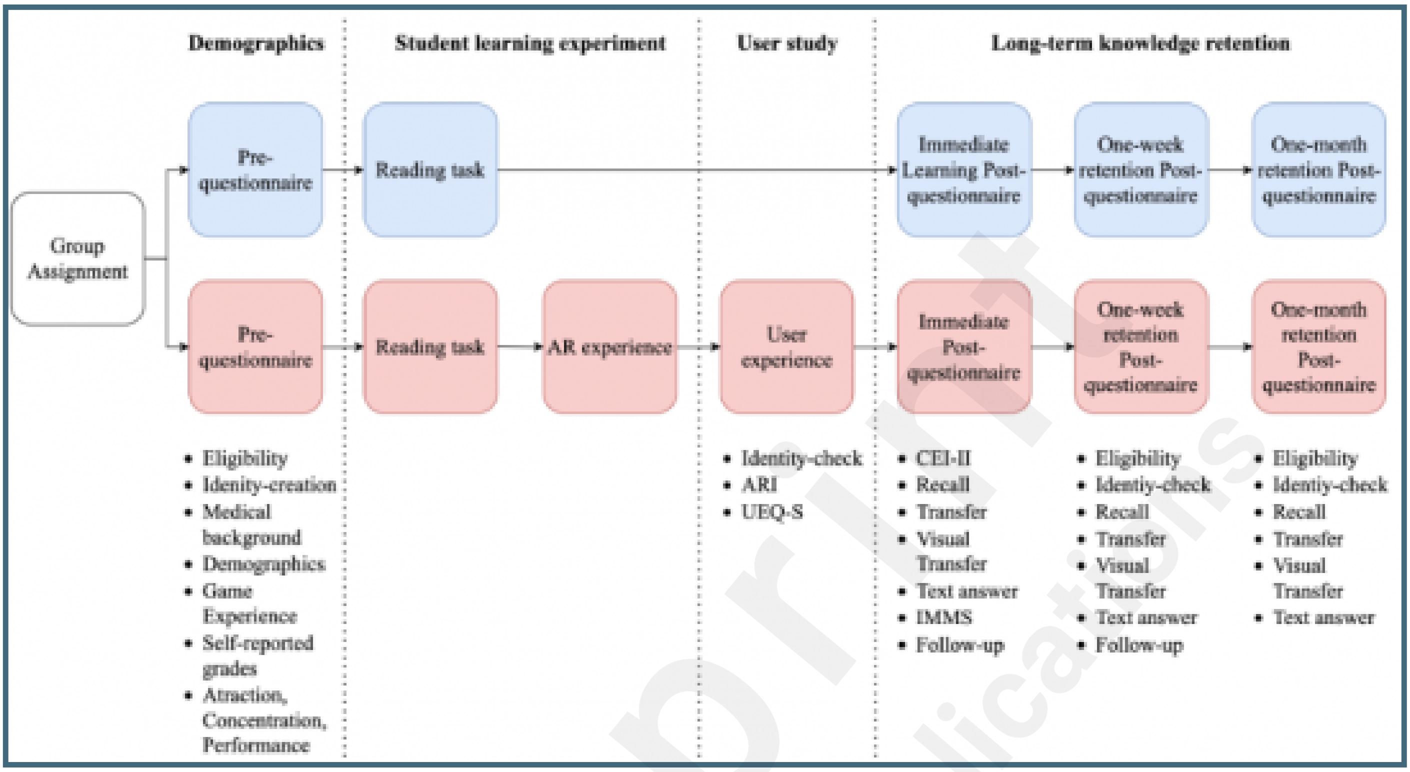
Cities. Remote Sens (Basel) MDPI AG; 2022 Jun 27;14(13):3095. doi: 10.3390/rs14133095
69. Chatain J, Kapur M, Sumner RW. Three Perspectives on Embodied Learning in Virtual Reality: Opportunities for Interaction Design. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems New York, NY, USA: Association for Computing Machinery; 2023. p. 1–8. doi: 10.1145/3544549.3585805
70. Bunt L, Greeff J, Taylor E. Enhancing Serious Game Design: Expert-Reviewed, Stakeholder-Centered Framework. JMIR Serious Games JMIR Serious Games; 2024 May 31;12(1):e48099. doi: 10.2196/48099
71. Metzmacher H, Rothnie F, Fayolle V, Sumner B, Zünd F. Using augmented reality games to communicate topics in Artificial Intelligence. 54th International Simulation and Gaming Conference (ISAGA 2023) ISAGA; 2023. p. 282–291. Available from: https://www.research-collection.ethz.ch/handle/20.500.11850/634829 [accessed May 31, 2024]
72. Colombo G, Grübel J. The Spatial Performance Assessment for Cognitive Evaluation (SPACE): A Novel Game for the Early Detection of Cognitive Impairment. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems New York, NY, USA: Association for Computing Machinery; 2023. p. 1–6. doi: 10.1145/3544549.3583828
73. Grübel J. The design, experiment, analyse, and reproduce principle for experimentation in virtual reality. Frontiers in Virtual Reality frontiersin.org; 2023;4. doi: 10.3389/frvir.2023.1069423
74. Schade C, Stagge A. Paper Prototyping as a Method for the Evaluation of Serious game Concepts. search.proquest.com; 2020. doi: 10.34190/GBL.20.089
75. Winn B, Heeter C. Resolving conflicts in educational game design through playtesting. Innovate: Journal of Online Education learntechlib.org; 2006 Dec 1;3:6. Available from: https://www.learntechlib.org/p/104284/

Supplementary Files

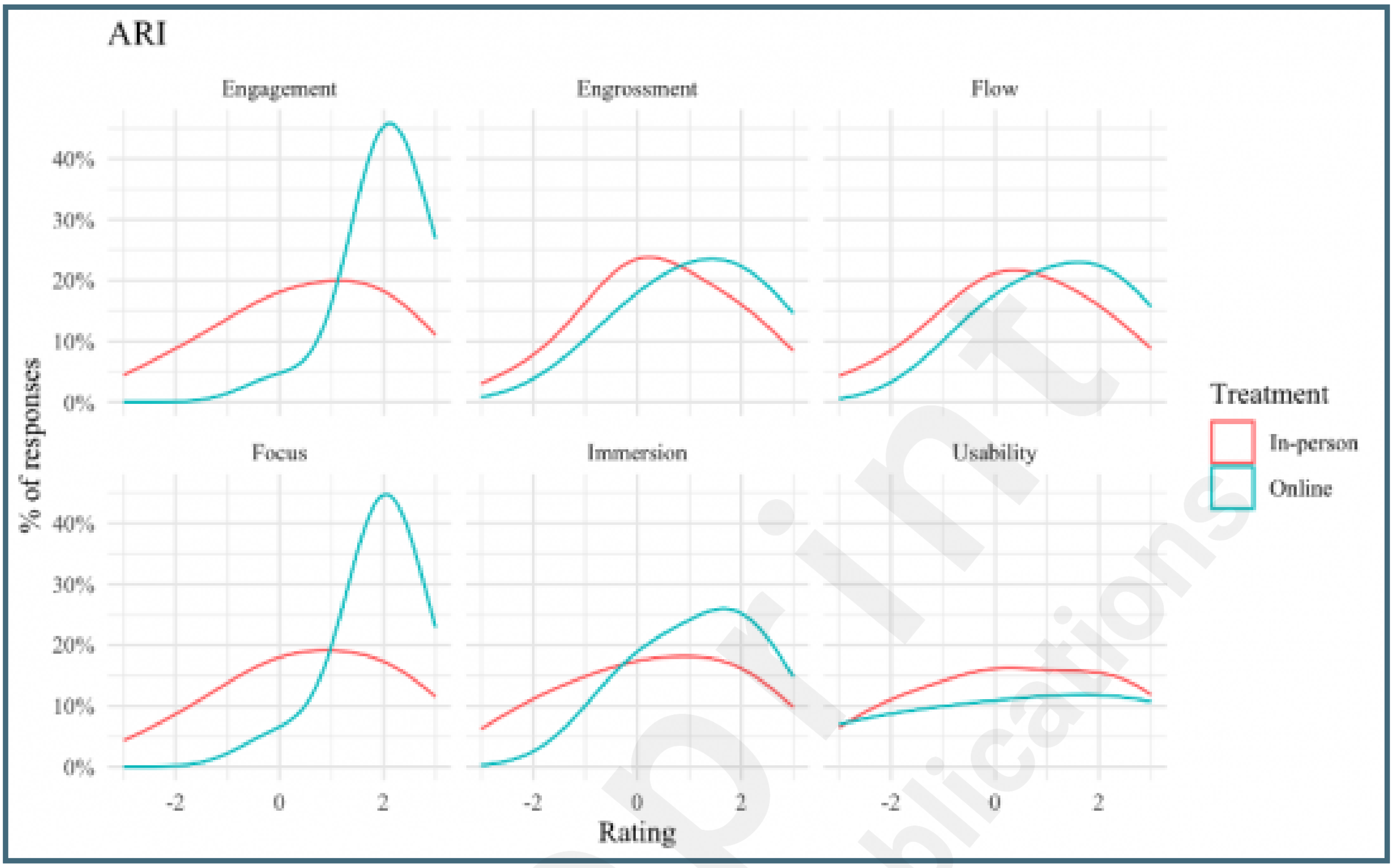
Figures

The AR Osteoclast serious game being played in a lecture hall.

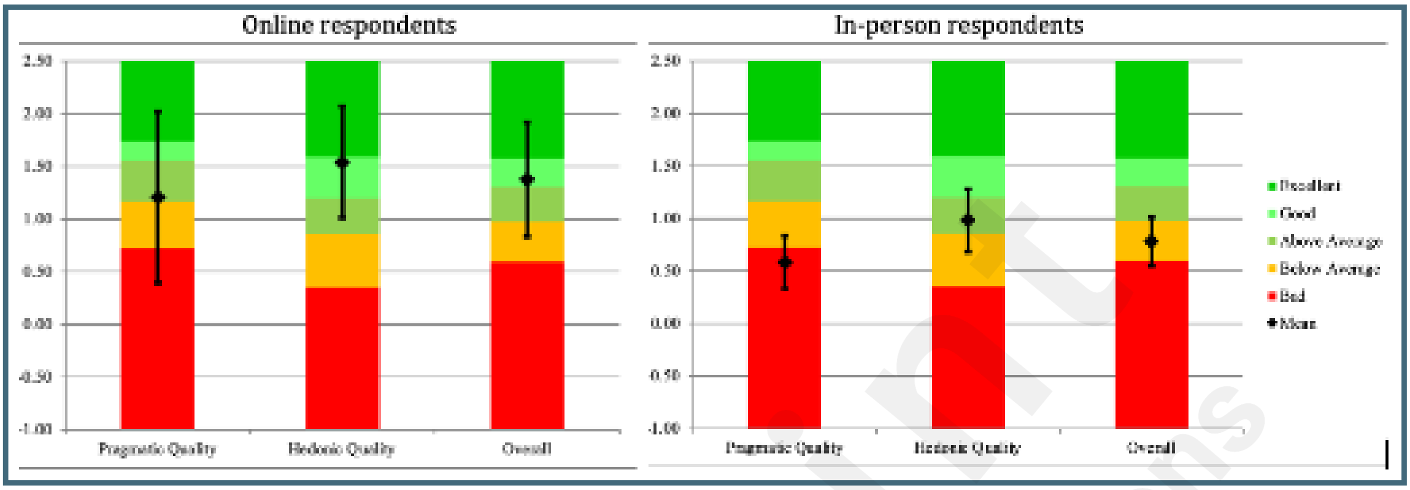
Protocol for the student learning experiment. Students are randomly assigned into the text-only group (blue) and the AR experience (red). General demographic information is acquired before the experiment and user study and the outcome is measured three times to gauge long-term retention of knowledge.



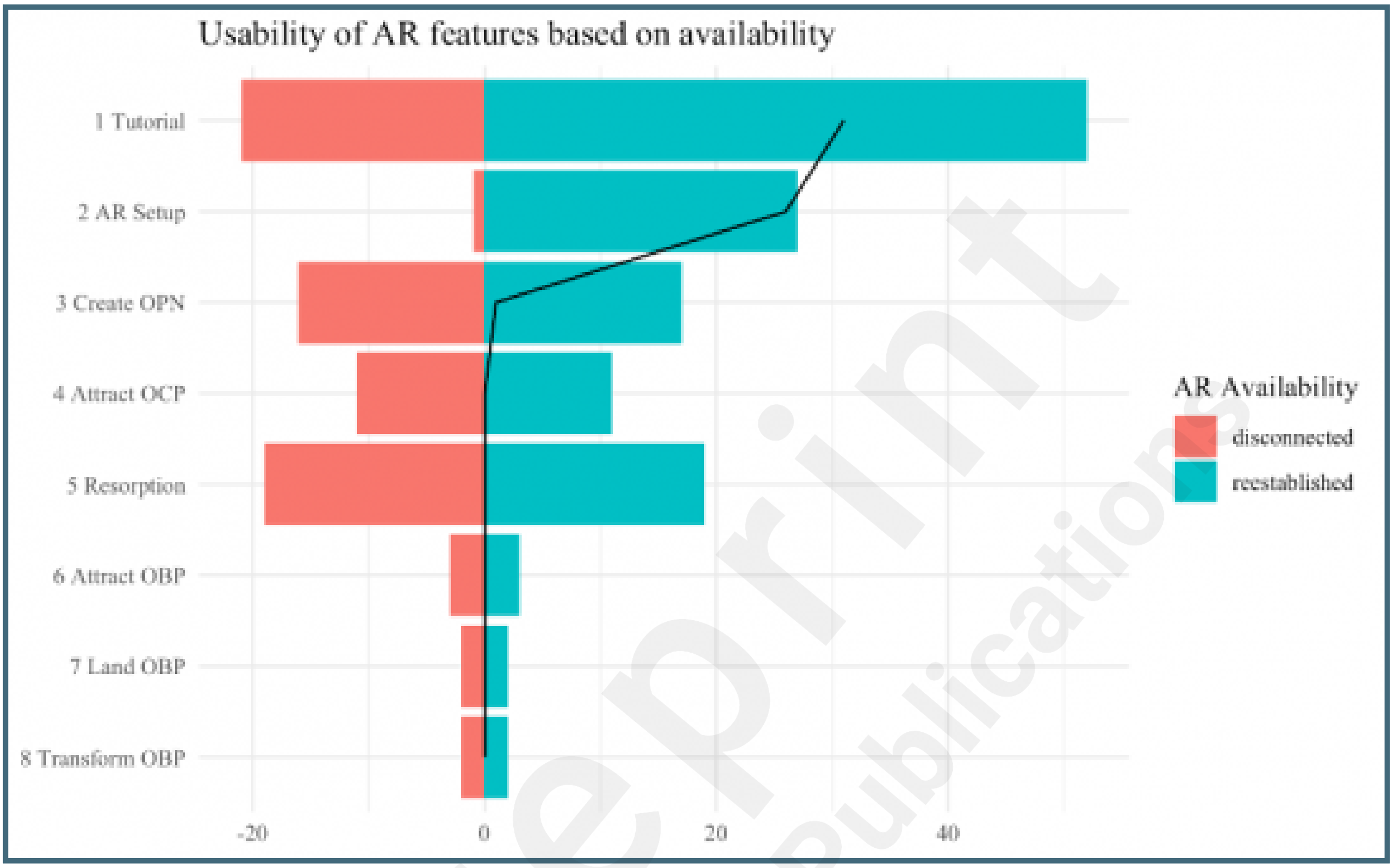
ARI results grouped by constructs from both the online participants and the student participants.



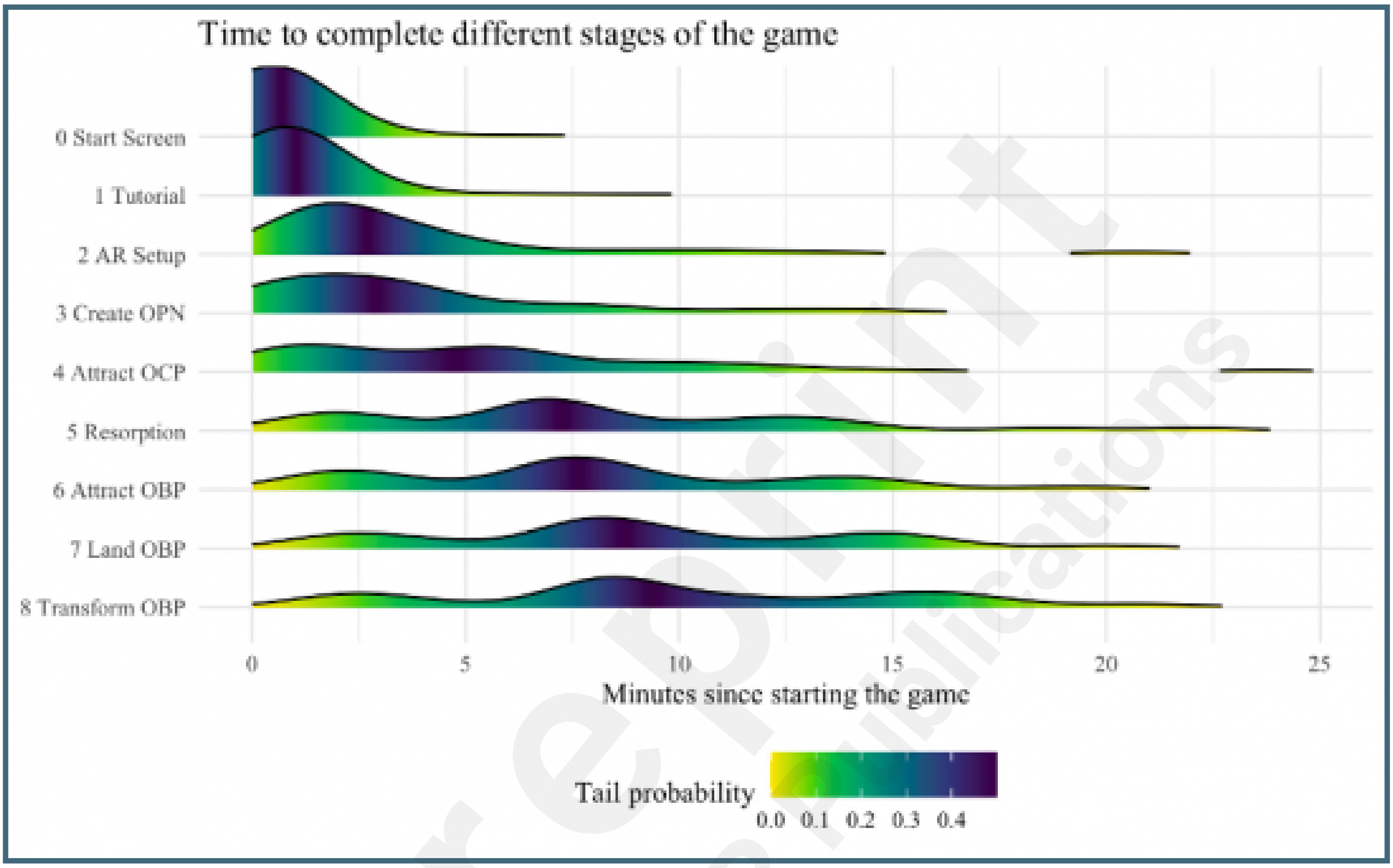
UEQ-S results from online respondents and in-person respondents. The online respondents have a more positive perception of AR osteoclasts placing its hedonic quality significantly above the mean of UEQ-S whereas in-person respondents only rate it slightly above average. The pragmatic quality is not significantly above average.



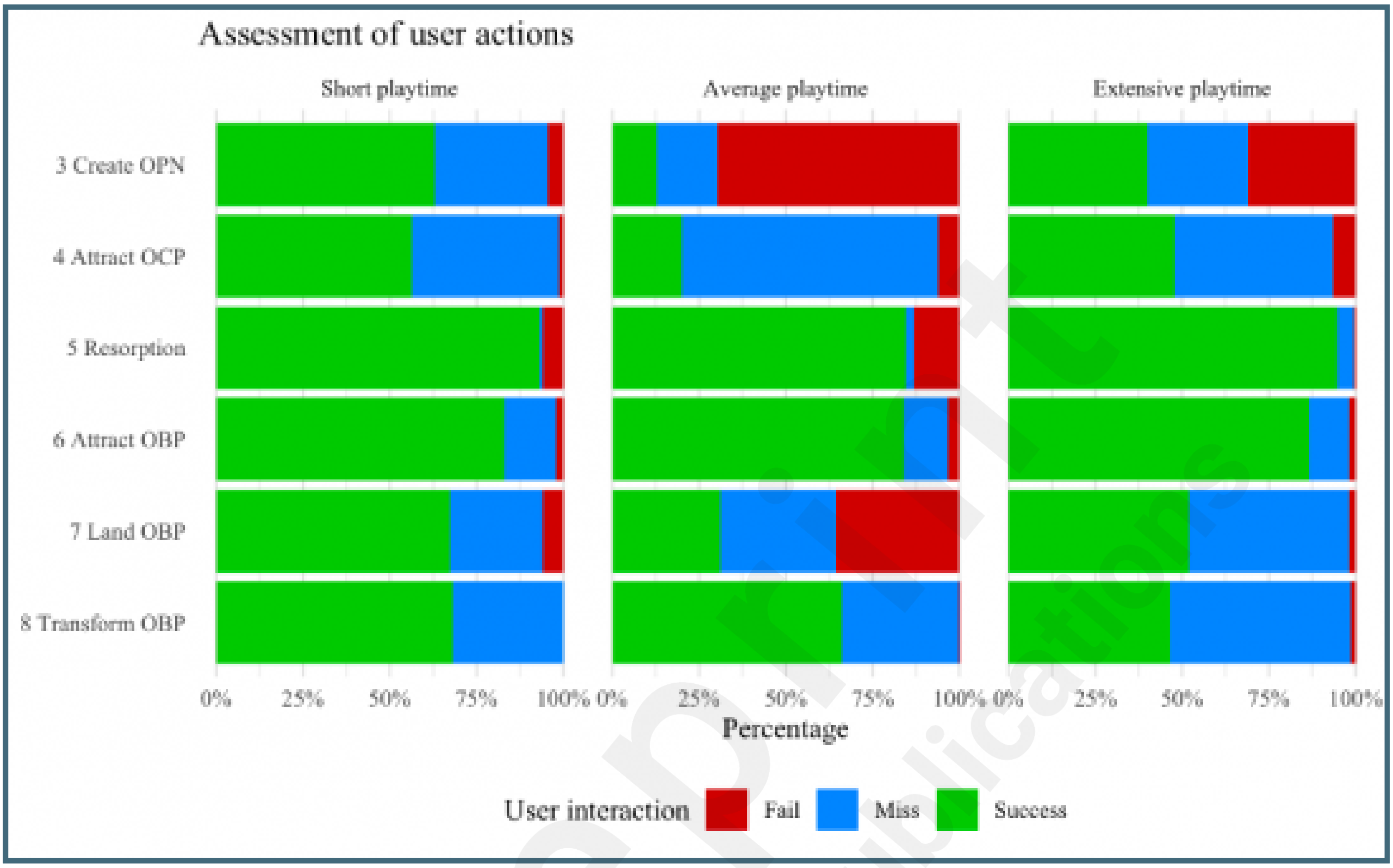
Usability of the AR features based on whether the AR was able to be displayed. Disconnected AR was caused when the camera could not see the book page. Connection was reestablished when the book page came into focus. During the Tutorial and the AR setup, most people acquire the AR display. During the game, the average is around 0 indicating that people were able to recover from losing the AR display.



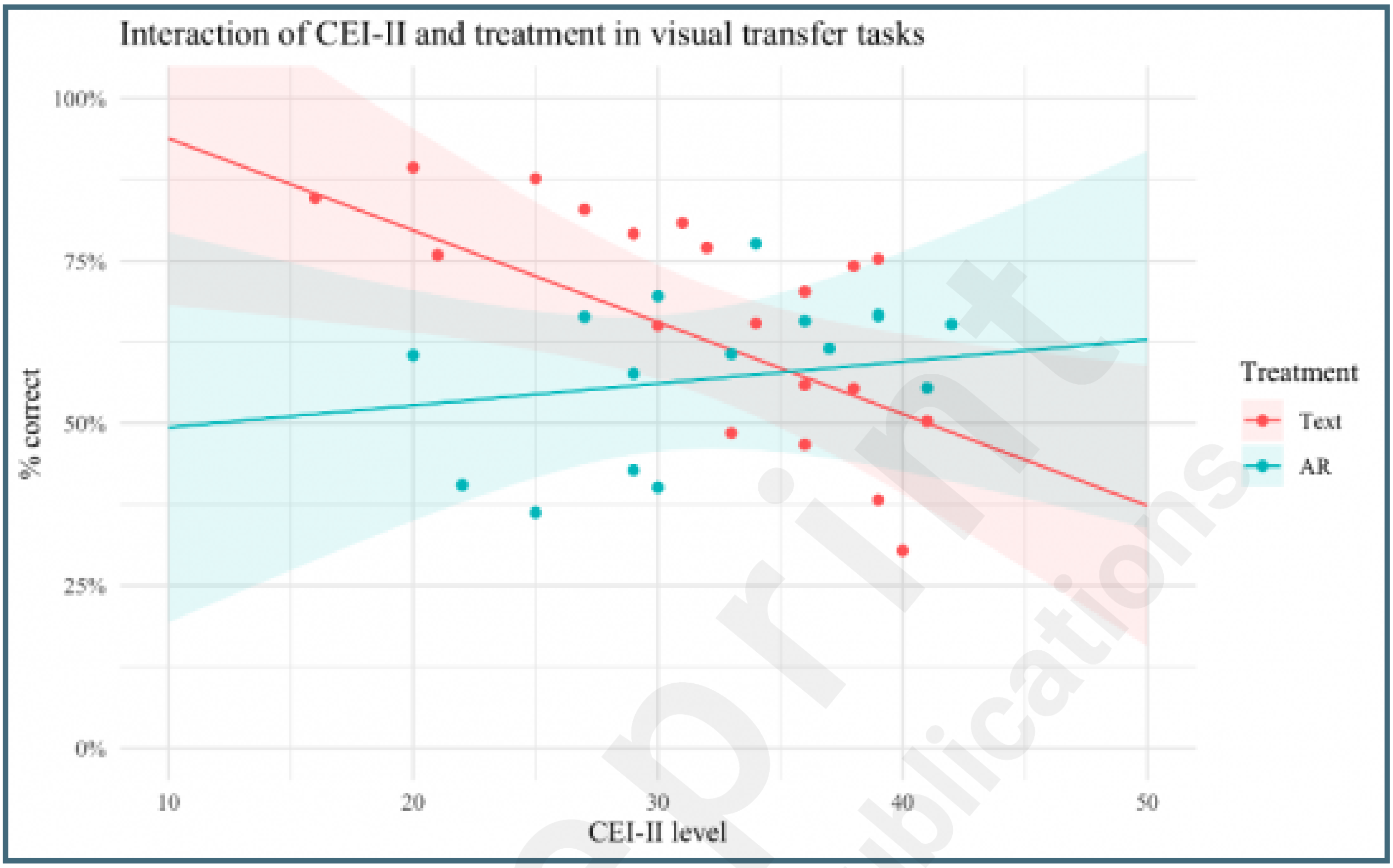
The stages of the game are shown from top to bottom. The time to complete is shown across all participants. The height of the ridges shows how many participants completed the stage at a specific minute. The colors are based on the empirical cumulative distribution function indicating that 95% of participants fall in the blue to green area and only 5% fall in the yellowish tones. We can see that the main cohort play the game in 8 minutes and that there are a few fast players completing it in 3 minutes and slow players completing the game in 16 minutes.



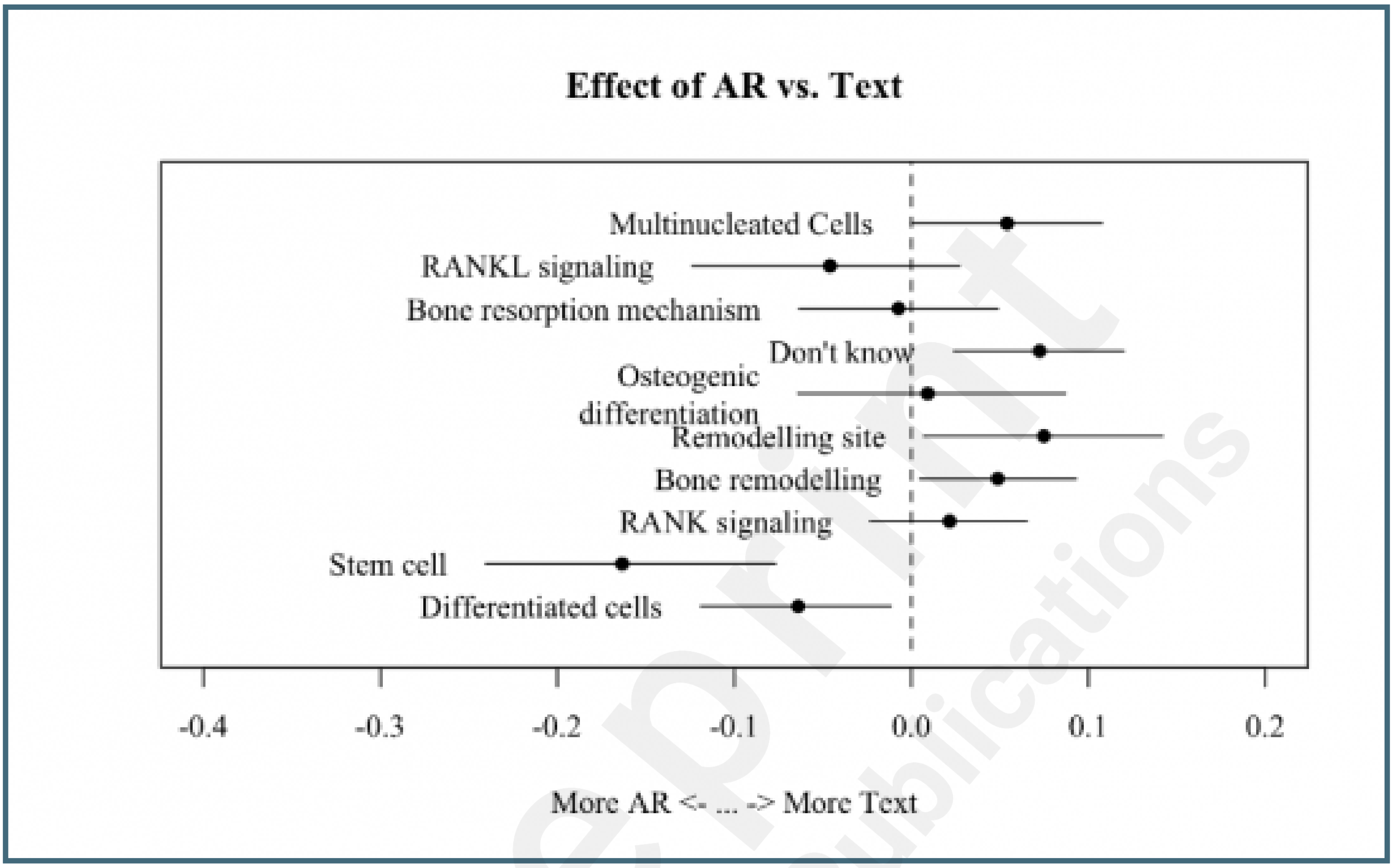
Short play times relate to higher success rates and extensive playtimes have higher miss rates. Most failures are occurring with average playtime, especially at the beginning of the game.

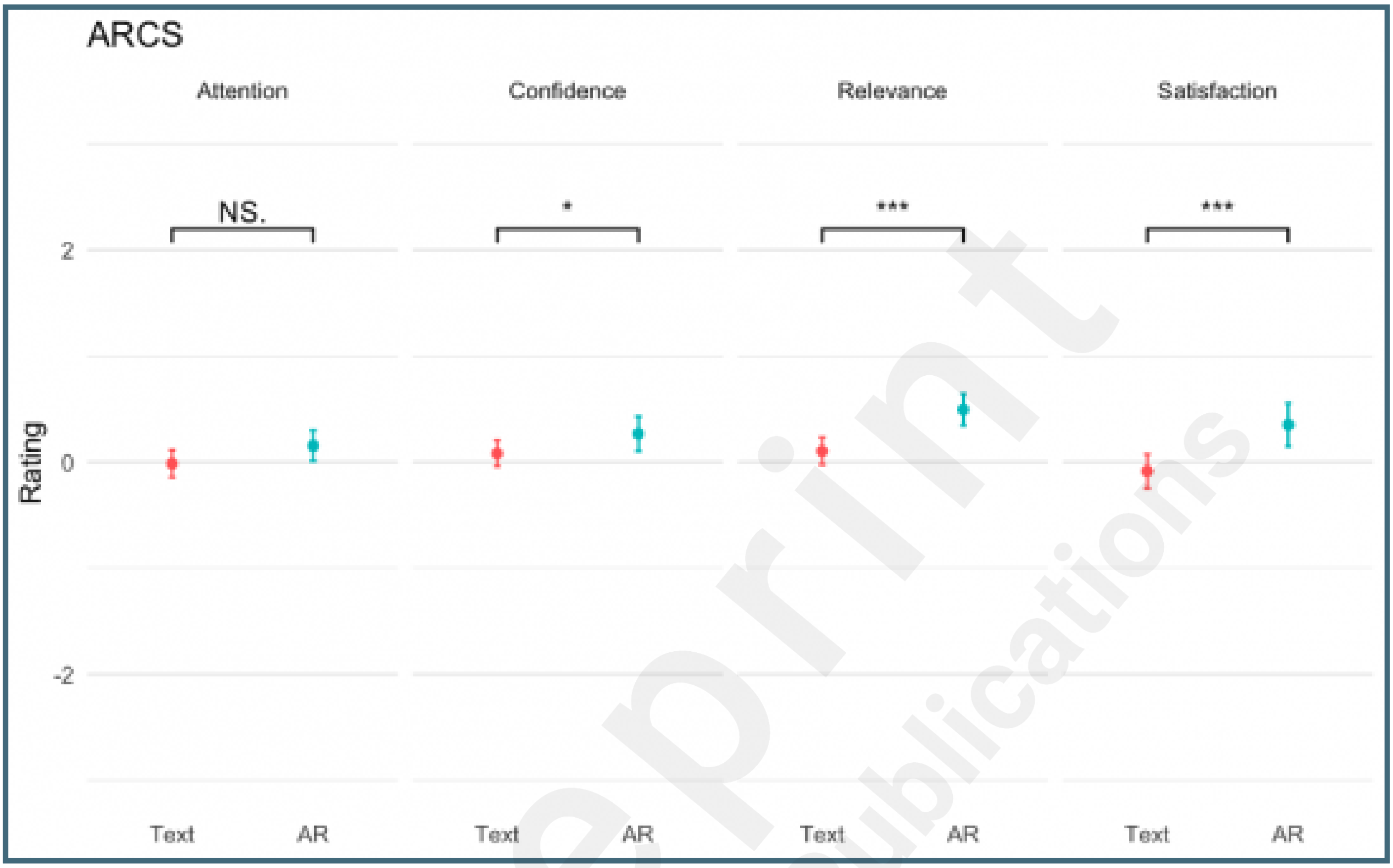


Our model reveals a strong inverse relationship between reading and curiosity in the students. For students that are curious, this can be offset by experiencing the information in AR.



Topic prevalence by the treatments. Negative values indicate a leaning towards the AR treatment, positive values indicate a leaning towards the text-only treatment. If the dashed line is touched, there is no significant leaning towards either side detected.




ARCS
Attention
Confidence
Relevance
Satisfaction
NS.
*
***
***
2
0
-2
Rating
Text
AR
Text
AR
Text
AR
Text
AR

Multimedia Appendixes

Statistical analysis details.

URL: http://asset.jmir.pub/assets/405193e49e6efcc7e9acefa00faff841.docx

Overview of appendices 3-5 and in-depth study of the game design.

URL: http://asset.jmir.pub/assets/0fc1f5f9f7378a605e40590ebebd78c7.docx

Video of playing AR Osteoclasts from beginning to end.

URL: http://asset.jmir.pub/assets/e197dce97c64455480657634da153918.mp4

Outlines of the survey questions fielded in Qualtrics.

URL: http://asset.jmir.pub/assets/5976f8e577dabb1a3fd48e1d222b68b4.docx

Zip file that contains survey data, text response data, and game data used in the analysis.

URL: http://asset.jmir.pub/assets/b6751261216182613a8c5e17147e4ef6.zip

TOC/Feature image for homepages

The AR Osteoclast serious game being played in a lecture hall.