Article

How to Evaluate Augmented Reality Embedded in Lesson Planning in Teacher Education

Anna Henne [1,2], Sabrina Syskowski [1,2], Manuel Krug [1,2], Philipp Möhrke [3], Lars-Jochen Thoms [1,2,*] and Johannes Huwer [1,2]

1 Science Education, University of Konstanz, 78464 Konstanz, Germany; anna.henne@uni-konstanz.de (A.H.); sabrina.syskowski@uni-konstanz.de (S.S.); manuel.krug@uni-konstanz.de (M.K.); johannes.huwer@uni-konstanz.de (J.H.)
2 Science Education, Thurgau University of Education, 8280 Kreuzlingen, Switzerland
3 Department of Physics, University of Konstanz, 78464 Konstanz, Germany; philipp.moehrke@uni-konstanz.de
* Correspondence: lars.thoms@phtg.ch

Abstract: Augmented reality (AR) is vital in education for enhancing learning and motivation through interactive environments and experiments. This requires teacher training in AR creation and integration. Research indicates that learning effectiveness relies on thorough preparation, calling for the development of scoring rubrics for evaluating both educational AR and AR's educational integration. However, no current studies provide such a rubric for assessing AR's pedagogical implementation. Hence, a scoring rubric, EVAR (Evaluating Augmented Reality in Education), was developed based on the framework for the analysis and development of augmented reality in science and engineering teaching by Czok and colleagues, and extended with core concepts of instructional design and lesson organization, featuring 18 items in five subscales rated on a four-point Likert scale. To evaluate the validity and reliability of the scoring rubric, AR learning scenarios, designed by eleven master's seminar pre-service teacher students at the University of Konstanz, majoring in biology, chemistry, or physics, were assessed by five AR experts using the newly developed scoring rubric. The results reveal that a simple classification of AR characteristics is insufficient for evaluating its pedagogical quality in learning scenarios. Instead, the newly developed scoring rubric for evaluating AR in educational settings showed high inter-rater reliability and can discriminate between different groups according to the educational quality of the AR and the implementation of AR into lesson planning.

Keywords: augmented reality; assessment; rubric; lesson planning; science education; teacher training

1. Introduction

Augmented reality (AR) is a technology that enables the real-time integration of virtual objects into our physical reality. Virtual objects can be placed in a real environment using various mobile devices, such as smartphones, tablets, or head-mounted displays (e.g., special AR glasses) [1]. Applications ranging from medicine to assistance with technical problems using AR are already common practice. In education, interactive learning environments can be created and various applications realized, from visualizing the invisible to expanding educational materials and enabling new experiments [2–4]. Given the pivotal role of schools in preparing students for the future, the integration of AR in educational settings is becoming increasingly important. Initial studies have shown positive effects of AR on motivation [5–11], self-efficacy [5,6], self-regulation [12], enjoyment in experimentation, attitude [5,6,13,14], and learning performance [5,6,14–29], especially in understanding abstract concepts [14–19,25,26,30–43] or including auxiliary information [3,24,44–46]. Furthermore, AR improves laboratory skills [13,47,48] and enables collaborative work [49]. Nevertheless, studies indicate that it is not intended to replace physical labs but that the experience of a virtual learning environment leads to an improved

understanding of, for example, chemical concepts [50]. Using AR usually contributes to a sufficient degree of good learning outcomes and is especially beneficial to long-lasting learning outcomes [51]. However, simply using AR does not guarantee learning success [51]. Hence, it is crucial to adapt teacher training to align with the requirements and potential that AR brings to an educational context [52–54]. Teachers should be equipped to create augmented reality experiences [55] and to select a suitable AR from existing options for their teaching scenarios [53,56]. The previous experience of pre-service teacher students gained during their university studies is hypothesized to have a positive influence on the actual use of media at school [57]. Nevertheless, only a few students come into contact with augmented reality during their studies [57]. Hence, it is important to offer a learning opportunity for this as part of the course.

Research in the realm of experimentation has revealed that the effectiveness of student learning depends on thorough preparation before and follow-up processing after the experimentation. Hence, it is crucial how activities are embedded in the lesson [58,59]. To assess the structured and didactically sound integration of AR into lessons or lesson plans, a scoring rubric is needed to allow for objective, reliable, valid, and test-economical measurement during lesson observations, including mock trials, for example. While research-based frameworks for the analysis and development of augmented reality in science and engineering teaching already exist [32,60], they need to be extended to incorporate aspects of instructional implementation, as this framework does not include any aspects of embedding experiments in lessons. However, we know from research on carrying out experiments that it is precisely the embedding of experiments in the classroom that is of particular importance for students' learning success. In addition to the categories from Czok et al. (namely *adaptivity*, *interactivity*, *immersion*, *congruence with reality*, *content proximity to reality*, *game elements*, and *complexity*), we also need aspects of instructional embedding such as "frictionless function of the AR", "confidence of the teacher when handling the AR", "simplicity of handling for the learners", "promotion of a learning objective through the AR", "embedding in the lesson", "design laws", and "cognitive load". Overall, these aspects can be divided thematically into the following main categories: "Technical Implementation", "Fit of the AR", "Interactivity and Engagement", "Visualization", and "Creativity and Originality". This leads to the following research questions:

1. Which categorizations, according to Czok et al. [32], can be found in augmented reality embedded in teaching scenarios created by pre-service teacher students in a master's seminar for teacher education?
2. How can the quality of the embedding of augmented reality in teaching be evaluated?
3. To what extent can the deductively derived structuring of the categorizations be mapped to reliable subscales?
4. To what extent does the quality of an AR learning environment determine the overall quality of the lesson planning integrating this AR learning environment?

2. Methods and Materials

2.1. Sample

Eleven pre-service teacher students (six female, five male; seven biology, four chemistry, and three physics students; multiple answers were possible, as in Germany teacher students select at least two subjects) voluntarily took part in this study during a master's seminar on science education at the University of Konstanz in the summer term of 2023. The participants were divided into six groups of one to two people (the students were allowed to choose their own partner).

2.2. Instrument

To assess the characteristics of augmented reality environments, the evaluation criteria proposed by [32] were used to answer research question one. A new scoring rubric to evaluate the use of augmented reality in teaching scenarios was developed to examine the second research question. The rubric is based on core concepts pronounced by Czok

et al. Based on the design parameters described there, it was checked for each category which form of lesson embedding is necessary to emphasize this aspect of learning with AR, and then a corresponding item was formulated. This was done for identifying ways teachers should use and embed AR in lessons, which are most beneficial. For example, in the area of interactivity [32], the contribution the quality of AR makes to the lessons was examined. It is not enough for the AR to be able to interact with the learners. Rather, the newly developed rubric should record whether it is actually being used in a meaningful way. The items were thematically grouped into five dimensions: *Technical Implementation*, *Fit of AR*, *Interactivity and Engagement*, *Visualization and Creativity*, and *Originality of AR Use*. A four-point Likert scale from "1—strongly agree" to "4—strongly disagree" was used. In addition, the option "no answer possible" could be selected. Notes could be added to each item to allow for a complementary qualitative assessment. Further, the rubric was consensually validated by six science education researchers with experience in the development and implementation of augmented reality in teaching and teacher training. The scoring rubric, EVAR (Evaluating Augmented Reality in Education), can be found in Table 1 and downloaded as supplementary material Table S1.

Table 1. The five subscales of the scoring rubric EVAR (Evaluating Augmented Reality in Education) with the corresponding 18 items. The terms teacher and learner are used in the scoring rubric. Teachers refer to those who have created or selected the augmented reality (including pre-service teachers or trainees). Learners here refer to those who act as participants in the teaching scenario (e.g., fellow students in the seminar context or pupils).

Subscale	Item	Item text
Technical Implementation	1	The AR in the learning scenario operates smoothly and reliably.
	2	The teachers are confident in controlling the AR.
	3	The handling of the AR is intuitive and simple for the learners.
	4	The functionality of the AR is sufficiently described and explained.
	5	The tracking method chosen is appropriate for the teaching scenario.
Fit of the AR	6	The AR supports at least one specific learning goal.
	7	There is a connection to previous and subsequent teaching sequences.
	8	Relevant references to real situations or applications are made.
	9	The AR offers clear benefits compared to conventional visualizations.
	10	Potential benefits and challenges of AR use for teaching are discussed.
	11	The AR helps learners to develop a better understanding of the content.
Interactivity and Engagement	12	The AR encourages learners to actively engage with the subject matter.
	13	There are additional possibilities besides viewing the object, e.g., interactivity or individualization.
	14	There are feedback mechanisms (analogous or digital) to provide learners with feedback on their use of AR.
Visualization	15	The complexity of the AR (cf. [60]) fits the learning goal addressed (in terms of cognitive load [61]).
	16	The design laws [62] are taken into account.
Creativity and Originality	17	The lesson design demonstrates an original and creative use of AR to support the learning process.
	18	The AR was created by the teachers themselves.

As part of the expert survey, items 7, 8, and 9 were shortened and simplified: (changes crossed out) "~~Integration into the teaching process~~: There is a connection to previous and subsequent teaching sequences." (I7). "~~Integration into the course of the lesson~~: Relevant references to real situations or applications are made." (I8). "~~Added Value and educational benefits~~: The AR offers clear benefits compared to conventional visualizations." (I9). Item 11 was assigned to "Fit of the AR"; before the expert survey, it was thematically assigned to "Visualization". Items 13 and 15 were supplemented with examples or references for precision (extension in brackets): "There are additional possibilities besides viewing the

object (e.g., interactivity or individualization.)" (I13) and "The complexity of the AR (cf. [62]) fits the learning goal addressed (in terms of cognitive load [63])." (I15). Finally, item 16 was newly added.

2.3. Study Design

Six groups of pre-service teachers presented self-created lessons in a 45-min presentation. Parts of the lessons were also carried out with fellow students in a mock trial. The subject areas were specified by the experts to ensure that the topic was fundamentally suitable for the use of AR. Five observing AR experts in the field of science education research participated during the presentation and applied the developed rubric EVAR (live and on site). Subsequently, Czok's questionnaire was applied to the submitted AR materials.

2.4. Context

The study was conducted during a master's seminar [63,64], especially targeted at the development of digital competencies for teaching in science education, and took place in the summer term of 2023 at the University of Konstanz. This seminar was divided into three phases (see Figure 1). In the first part, the basics of teaching with digital media were introduced and practiced in alternating theory and voluntary on-site exercises with a team of tutors following the DiKoLAN framework [65,66]. DiKoLAN is a competency framework that defines seven digital core competency areas that science education students should have acquired by the end of their studies.

Figure 1. The structure of the seminar with its three phases: initial lecture phase, second project phase, and final presentation.

DiKoLAN is divided into two sections: general competency areas, encompassing *documentation*, *presentation*, *communication/collaboration*, and *information search and evaluation*, as well as competency areas specific to the natural sciences, including *data acquisition*, *data processing*, and *simulation and modeling*. In the context of this seminar, particular emphasis was placed on the application of AR as an example of simulation and modeling, as well as the creation of AR content [64]. Here, students acquire essential knowledge about models, their development, and the application of AR. AR is explained through the lens of seven design parameters, according to [32]. Each parameter has different levels or indicators that enable a comparison between different AR implementations. These parameters are *adaptivity*, *interactivity*, *immersion*, *congruence with reality*, *content proximity to reality*, *game elements*, and *complexity*. Adaptivity describes the program's ability to adjust to various situations by reacting to activities, events, or changes in situations. Interactivity refers to the intended interaction between the user and the digital media components and includes six levels of interaction. According to [32], immersion is understood as the ability of digital media to influence human senses, and the degree of immersion increases as more

senses are engaged. Congruence with reality assesses the plausibility and realism of AR implementations in terms of social and perceptual realism. Content proximity to reality examines the plausibility of AR content regarding causal, spatial, and temporal factors, as well as the tracking method's appropriate use. The incorporation of game elements in education can enhance interactivity and motivation. For this parameter, eight indicators are provided. Lastly, complexity reflects the content-related and cognitive structures of AR functions, whereby achieving a higher level of complexity is associated with a higher demand on the user or more extensive cognitive activity. In general, the aim of the seminar is to promote the future-oriented and didactically sound use of digital tools in science lessons. Therefore, students are trained to create AR content themselves and analyze suitable tools. Hence, AR is not viewed merely as a technical gimmick but as a powerful tool for future educators, which could change teaching science in school.

In the second part of the seminar, students planned teaching sequences for upper-level classes in groups of two, incorporating AR. The prospective teachers were given a predetermined topic from the field of molecular orbital theory, including core chemistry, underlying physics concepts, and biological contexts. The students could choose one of these secondary 2 / undergraduate science topics for which AR visualization is promising. It was, therefore, not a matter of investigating whether AR makes sense in general. It can be assumed that the prerequisites for (meaningful) AR are given in principle. Clear guidelines are provided by the instructors after diagnosing the potential of AR. The students had around four weeks to plan a teaching unit, select or develop an AR, and implement it into a teaching unit (i.e., a lesson plan). To enhance the clarity and understanding of the results, the created teaching sequences are presented in Table 2. The educational offerings included lectures, exercises, and DiKoLAN sessions with individual supervision [63]. Additionally, the opportunity to enhance their education was offered through self-learning units on the DiBaNa website (DiBaNa: Digitale Basiskompetenzen in den Naturwissenschaften, German for Digital Basic Competencies for Science Teachers), an online platform for acquiring digital teaching competencies [67].

Table 2. Topics of teaching sequences, learning goals, and AR implementation by group.

Group	Topic	Learning Goals	AR
1	Construction of Amino Acids	Learners determine the concept of chirality through the structure of amino acids.	AR model of *L-alanine* for comparison with the model built using the model kit (Tools: TinkerCAD [68] and Zapworks Designer [69])
2	Properties of Enzymes Illustrated Using the Bioluminescence of Fireflies	Learners describe the structure and properties of an enzyme, explaining its mechanism using appropriate models: the key–lock principle and induced fit model.	AR model of *key–lock principle* (TinkerCAD [68] and Zapworks Designer [69])
3	Introduction to Chirality	Learners can explain the chirality of a molecule based on the presence of an asymmetrically substituted carbon atom.	Aligning AR models of *chiral or achiral substances* with the model kit (TinkerCAD [68] and zapworks designer [69])
4	Introduction to Orbital Theory	Learners recognize the relationship between a wave function and an orbital representation	AR model of the *orbital representation* (Geogebra [70])
5	Bond Formation by Orbital Overlap	Learners apply their knowledge of orbital models to simple atomic bonds.	AR model of the *molecules and first simple connection* (leARnchem [71])
6	Superposition Principle	Learners can explain wave phenomena such as the path difference and constructive and destructive interference with the superposition principle.	AR simulation of the *interference pattern of transversal waves with two emitters* (Geogebra [70])

In the third phase, each group of pre-service teachers presented their planned lessons, and a written elaboration was handed in.

2.5. Statistical Analysis

For many of the items, a bimodal distribution of responses was expected (separating agreement and disagreement). In addition, all items were positively worded so that students would desirably achieve a positive rating with their unit after the training measure. This leads to the expectation of a one-sided distribution. In the evaluation, therefore, statistical measures such as Cohen's kappa [72], Fleiss' kappa [73], or Krippendorff's alpha [74] were not applicable since these are known to be problematic with a one-sided distribution, especially for small sample sizes [75]. Instead, a graphical method was used for the evaluation of the inter-rater reliability, and a frequency map was created with the statistical software R [76]. In order to compare the results of the two parts of the research, the mean value across all seven areas according to [32] was compared with the mean value across all 18 self-generated items for each group using Excel [77]. To check the scale reliability, Guttman's lambda 4 and lambda 6 [78] were calculated with R. Cronbach's alpha [79] was unsuitable since no normal distribution was expected due to the assumed polarization of the responses.

3. Results

3.1. Characteristics of the AR Used

For each augmented reality presented by the students, a classification of characteristic features was performed based on [32,60]. The results of the seven categories for all six groups were plotted on a spider web plot. As the different categories have different maximums of scores, the values reached in each category were normalized. The results are shown in Figure 2.

When looking over the spider webs, it is clear that for the implementation of each group, there are differences in which category is the most pronounced and whether the degree of pronouncement is rather the same across all categories or rather fluctuating. Only the groups represented by Figure 2c,d show the same values over all categories. In general, as a common feature, it can be found that all of the featured AR have low values for game elements (GE) as well as low immersion (Imm) values. For five out of six, the relative value for congruence with reality (CwR) is about 0.6; still, two-thirds of the apps have a relative value of 0.4 in terms of content proximity to reality (CPtR), and half of them have the same values for complexity (Comp) and interactivity (Int), but for the other categories, the spider webs show different figures.

3.2. Evaluation of the Teaching Scenarios including AR

The aim is to verify whether the newly developed scoring rubric can serve as an evaluation basis for the instructional context in which augmented reality was used. For this purpose, the inter-rater reliability is presented below.

For the representations of groups 3, 5, and 6, the raters ranked the majority of items very similarly high (cf. Figure 3). For groups 1 and 4, the scatter is wider, while for group 2, the scatter of the ratings is the widest. The variance in the scatter can be easily recognized in the heat map.

In addition to the significant differences between the groups, it is also striking that for items 4, 10, 13, and 14, the answers are distributed over three to four neighboring scale levels, or there is contrasting checkbox behavior for at least three groups, indicating a low level of agreement among the raters.
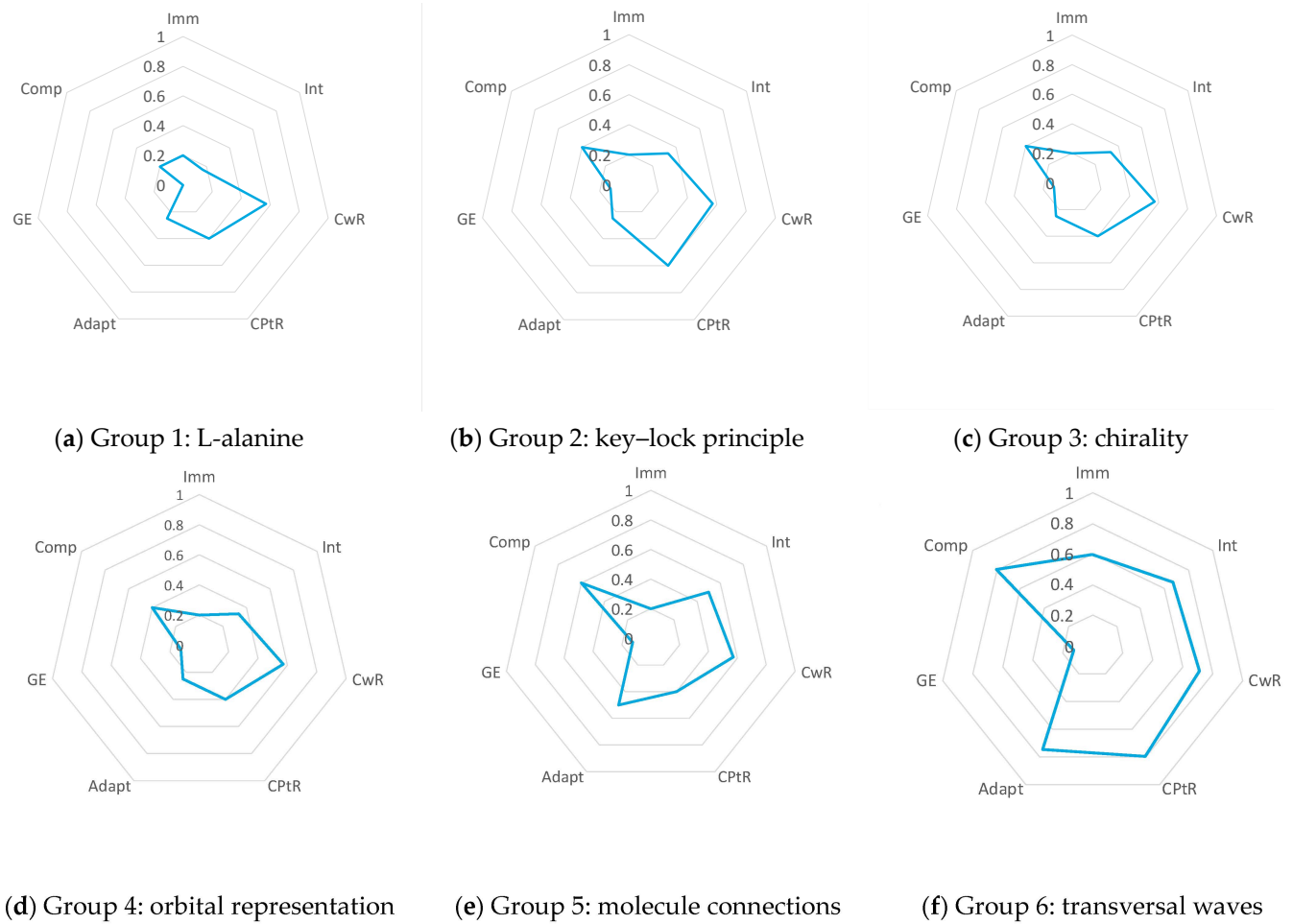
(**a**) Group 1: L-alanine

(**b**) Group 2: key–lock principle

(**c**) Group 3: chirality

(**d**) Group 4: orbital representation

(**e**) Group 5: molecule connections

(**f**) Group 6: transversal waves

Figure 2. The results of the classifications according to [32] from (**a**) group 1 to (**f**) group 6 in the seven categories: adaptivity (Adapt), interactivity (Int), immersion (Imm), congruence with reality (CwR), content proximity to reality (CPtR), game elements (GE), and complexity (Comp).
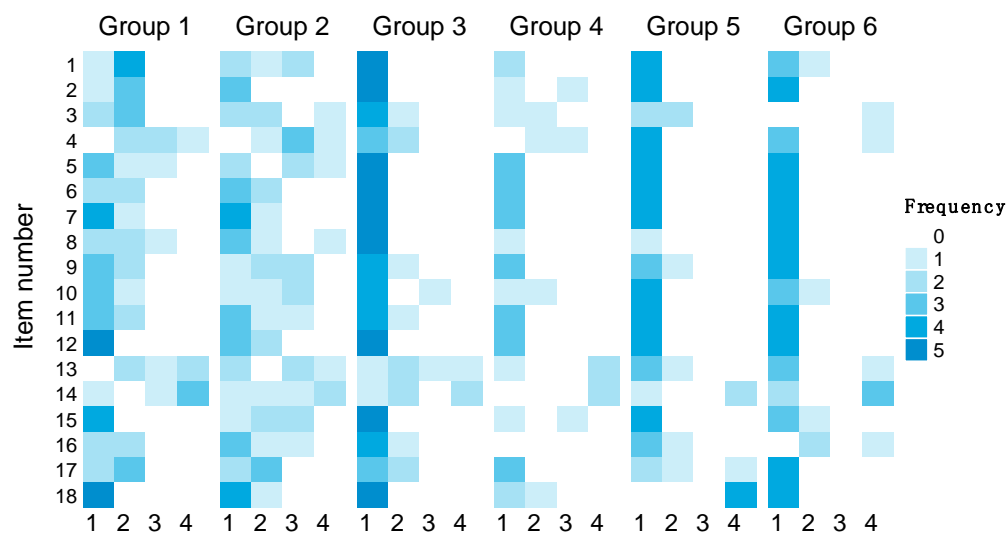

Figure 3. The frequency of the given answers broken down by answer options 1 ("I agree absolutely") to 4 ("I do not agree at all") across all 18 items.

3.3. Reliability of the Rubric and Its Theoretically Derived Subscales

Based on the concepts from Czok [32], the scale was divided thematically into subscales. To check the meaningfulness of this division, Guttman's lambda-4 and lambda-6 were calculated for the main scale and the five dimensions *Technical Implementation*, *Fit of the AR*, *Interactivity and Engagement*, *Visualization,* and *Creativity and Originality*. The results are illustrated in Table 3. Further, the absolute frequencies of the individual response options for each item across all groups and the relative frequencies are shown. For the main scale, a lambda-4 of 0.93 and a lambda-6 of 0.99 were reached.

Table 3. The counted answers for each item across all groups as absolute values and as relative values. The last columns contain the calculated values for lambda-4 and lambda-6 for each subscale.

Subscale	Item	Item text	1	2	3	4	NA	1	2	3	4	NA	λ_6	λ_4
Technical Implementation	1	The AR in the learning scenario operates smoothly and reliably.	22	4	3	0	1	73%	13%	10%	0%	3%	0.85	0.89
	2	The teachers are confident in controlling the AR.	21	4	2	0	3	70%	13%	7%	0%	10%		
	3	The handling of the AR is intuitive and simple for the learners.	16	9	1	0	4	53%	30%	3%	0%	13%		
	4	The functionality of the AR is sufficiently described and explained.	16	6	6	2	0	53%	20%	20%	7%	0%		
	5	The tracking method chosen is appropriate for the teaching scenario.	23	3	2	1	1	77%	10%	7%	3%	3%		
Fit of the AR	6	The AR supports at least one specific learning goal.	25	5	0	0	0	83%	17%	0%	0%	0%	0.88	0.92
	7	There is a connection to previous and subsequent teaching sequences.	27	3	0	0	0	90%	10%	0%	0%	0%		
	8	Relevant references to real situations or applications are made.	19	5	1	0	5	63%	17%	3%	0%	17%		
	9	The AR offers clear benefits compared to conventional visualizations.	23	5	2	0	0	77%	17%	7%	0%	0%		
	10	Potential benefits and challenges of AR use for teaching are discussed.	19	5	4	0	2	63%	17%	13%	0%	7%		
	11	The AR helps learners to develop a better understanding of the content.	24	5	1	0	0	80%	17%	3%	0%	0%		
Interactivity and Engagement	12	The AR encourages learners to actively engage with the subject matter.	26	2	1	1	0	87%	7%	3%	3%	0%	0.48	0.67
	13	There are additional possibilities besides viewing the object, e.g., interactivity or individualization.	16	1	4	8	1	53%	3%	13%	27%	3%		
	14	There are feedback mechanisms (analogous or digital) to provide learners with feedback on their use of AR.	5	5	0	17	3	17%	17%	0%	57%	10%		
Visualization	15	The complexity of the AR fits the learning goal addressed (in terms of cognitive load).	20	7	3	0	0	67%	23%	10%	0%	0%	0.03	0.07
	16	The design laws are taken into account.	10	12	1	0	7	33%	40%	3%	0%	23%		
Creativity and Originality	17	The lesson design demonstrates an original and creative use of AR to support the learning process.	17	10	0	3	0	57%	33%	0%	10%	0%	0.67	0.80
	18	The AR was created by the teachers themselves.	21	2	0	5	2	70%	7%	0%	17%	7%		

3.4. Relevance of the Quality of an AR Learning Environment for the Overall Quality of the Lesson Planning

For further consideration, the mean value of each group with regard to the classification in all seven areas according to [32] was compared with the mean value of each

group with regard to the classification in all seven areas of the developed rubric. The results are shown in Figure 4. A slight correlation can be seen between achieving a high score according to [32] and a higher evaluation of teaching commitment. Around 19.4% of the variance in the evaluation of lesson planning can be explained by the quality of the integrated AR learning environment.
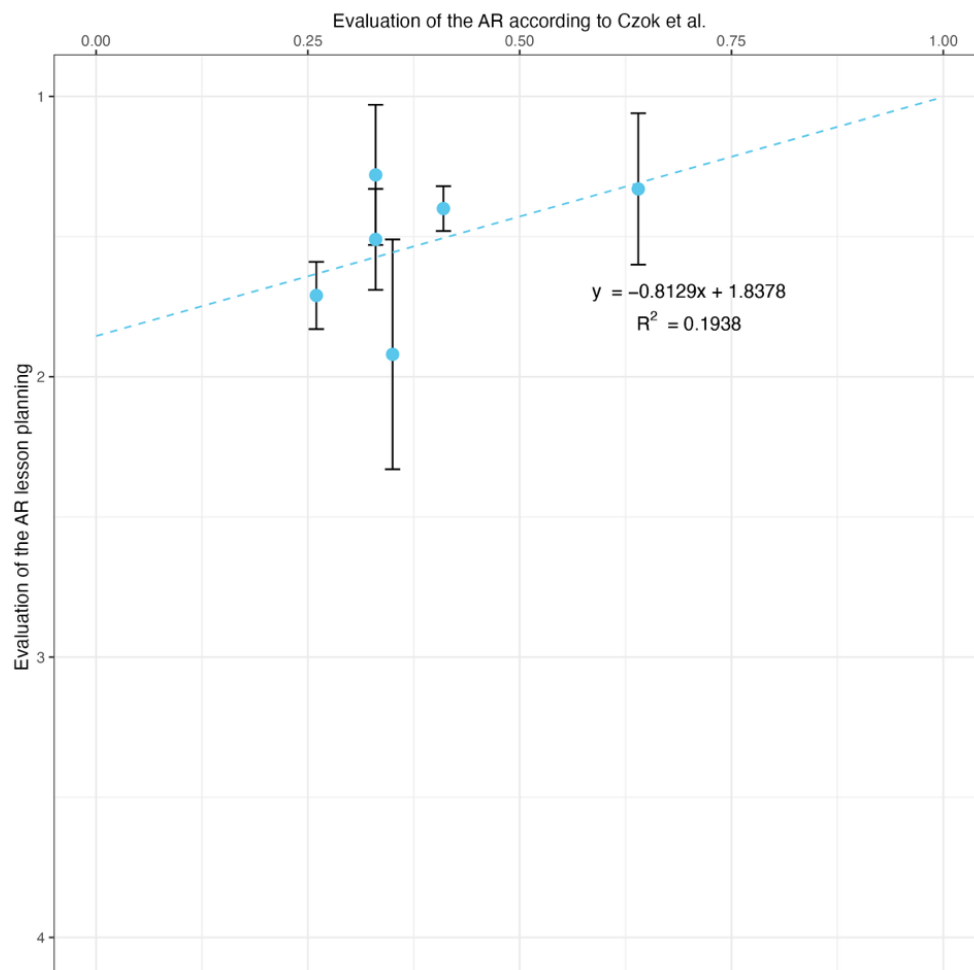

Figure 4. The comparison of the assessment from the two research parts. In the evaluation according to [32], 0 stands for no points achieved and 1 for maximum points achieved in all seven areas. In the rubric presented here, 1 stands for "I totally agree" and 4 for "I totally disagree".

4. Discussion

4.1. Characteristics of the AR Used

As a common feature, it can be seen that low immersion values are often achieved. Furthermore, all ARs contain only a few or even no game elements. Apart from that, no specific preferred trend can be identified for the other areas. This clearly reflects the seminar content. On the one hand, only the immersion of further senses except the optical one was treated in the seminar. Likewise, game elements played no role in the design of the seminar unit on AR and therefore are not found in the environments designed by the students. On the other hand, no specific training was conducted with the aim of achieving particularly high scale levels in the other categories, which explains well the different high values for these categories.

The AR of group 5 stands out with high or higher item ratings in all areas. This can be attributed to the fact that this group used a very comprehensive augmented reality application, leARnCHEM [80], developed at the University of Toronto that already included many features, e.g., for individualization and different levels of complexity. In contrast, the

ARs of groups 1 to 4 focused on a specific problem that was needed in the teaching setting being worked on. For this, no more functions than necessary were included.

4.2. Evaluation of the Teaching Scenarios including AR

To verify that the rubric created is a way to evaluate AR in an instructional context, the inter-rater reliability was first examined. Basically, for the relatively high number of five raters in the six groups, a high level of agreement could be achieved for many of the items (see Figure 3). The different results for the different groups seem to allow a conclusion to be drawn about the technical quality of the AR used. While the AR worked very well for groups 3, 5, and 6, the realization was technically rather challenging for group 2. Nonetheless, group 2 delivered a very convincing and thoughtful instructional concept for the use of their AR in the classroom. Further research is needed to confirm this assumption. In addition, the students' challenges in the technical realization of AR seemed to make the ratings more ambiguous. This had to be addressed by making the raters aware of the issue and finding ways to address it. It may also be beneficial to provide advanced training to experts who offer advice, helping them determine whether to prioritize the technical implementation or the quality of the idea, particularly in cases where a good idea is compromised by poor technical execution.

Summarizing the results of the frequency map (Figure 3), items 4 ("The functionality of the AR is sufficiently described and explained."), 10 ("Potential benefits and challenges of AR use for teaching are discussed."), 13 ("There are additional possibilities besides viewing the object, for example, interactivity or individualization", and 14 ("Feedback mechanisms (analogous or digital) are in place to provide learners with feedback on their use of AR.") were assessed differently by the raters. A possibility to counteract this would be a more detailed coding guide to train the rating experts.

Romano et al. [56] have already called for the creation of augmented reality by preservice teachers, and several studies have discussed the advantages of using them in teaching scenarios [5–9,15–19,31,53–55]. However, none of these studies included a scoring rubric to determine the pedagogical quality of the augmented reality implementation.

4.3. Reliability of the Rubric and Its Theoretically Derived Subscales

The calculated lambda-6 values for the *technical implementation* and *fit of the AR* subscales are between 0.8 and 0.9 and thus show that the subscales formed on the basis of theory can be very well implemented. A still acceptable value is achieved for the *interactivity and engagement* subscales. It is only the combination of items 15 and 16 on the *visualization* subscale that cannot be confirmed by the calculation of Guttman's lambda. However, when looking at the cross-tabulation of the answers given, this summary cannot be refuted either (see Table 4).

Table 4. Comparison of the answers for items 15 and 16.

Item		V15						
		1	2	3	4	NA	valid	total
V16	1	7	1	2	0	0	*10*	10
	2	7	5	0	0	0	*12*	12
	3	0	1	0	0	0	*1*	1
	4	0	0	0	0	0	*0*	0
	NA	6	0	1	0			
	valid	*14*	*7*	*2*	*0*			
	total	20	7	3	0			

4.4. Relevance of the Quality of an AR Learning Environment for the Overall Quality of the Lesson Planning

Even though a slight connection between the two different ratings, the evaluation according to [32] and with the instrument EVAR, was found, it is clear to see that evaluating the design options according to [32] is an important factor but not a reliable indicator of good teaching embedding. This clearly requires a specific rubric for evaluating lesson embedding. In groups 3 and 4, for example, which were assessed in the same way according to [32], the assessment of AR in teaching embedding differs noticeably. This clearly shows why a pure classification based on different characteristics is not sufficient for the evaluation of AR in a teaching context since neither the occurrence of many different characteristics nor a focus on a few characteristics is better or worse without a teaching context.

4.5. Limitations

This study focuses on the validation of the rubric. Therefore, based on the available data, no statement can be made about the occurrence of certain characteristics of the population due to the small sample. A larger sample is necessary for this.

Considering the agreement between the raters, a subjective interpretation cannot be ruled out. The evaluation of the selected AR application was carried out individually during each presentation and the introduction of each lesson and was not revised after the end of each presentation. Therefore, finding a consensus among the raters can be ruled out, as the raters did not discuss the material.

The rubric was developed in order to evaluate the use of AR in the classroom for given topics that were classified by a team of experts as beneficial for the use of AR. Therefore, no items can be found that assess whether the selected topics are suitable for the use of AR at all.

5. Conclusions

EVAR fills a gap that gives teacher educators a tool that can be used to evaluate a teaching sequence on augmented reality. It offers the possibility of adding an assessment to some of the features beyond just looking at them. The use of the grid is useful when selecting an AR to determine whether it is conducive to the teaching purpose. By guiding the questions that the teacher has to ask the AR, the selection of the AR is made easier. The evaluation grid was further developed to provide teachers with an instrument that they can use to make a reflective decision as to whether creating a tool is worthwhile or whether an already-created tool or another alternative can fulfil the same learning objective. For example, it assesses whether the additional work involved in creating the tool could be worthwhile because it reduces the workload elsewhere, whether the desired learning objective can be achieved, or whether another alternative can achieve the same learning objective with less effort. It was shown that around 20% of the responses were already predicted in terms of the quality of the augmented reality. There is therefore an opportunity to carry out further investigations with an even larger sample, for example, to determine the influence of individual design parameters according to Czok on the individual subscales of EVAR.

6. Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used DeepL (www.deepl.com, accessed on 29 January 2024) and Grammarly (www.grammarly.com, accessed on 29 January 2024) in order to improve the readability and language of single sentences. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/educsci14030264/s1: Table S1: Scoring rubric, EVAR.

References

1. Cao, J.; Lam, K.-Y.; Lee, L.-H.; Liu, X.; Hui, P.; Su, X. Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence. *ACM Comput. Surv.* **2023**, *55*, 1–36. [CrossRef]
2. Gomollón-Bel, F. IUPAC Top Ten Emerging Technologies in Chemistry 2022. *Chem. Int.* **2022**, *44*, 4–13. [CrossRef]
3. Milgram, P.; Takemura, H.; Utsumi, A.; Kishino, F. Augmented reality: A class of displays on the reality-virtuality continuum. In Proceedings of the Photonics for Industrial Applications; SPIE: Bellingham, WA, USA, 1995; pp. 282–292.
4. Azuma, R.T. A survey of augmented reality. *Presence-Virtual Augment. Real.* **1997**, *6*, 355–385. [CrossRef]
5. Cai, S.; Chiang, F.-K.; Sun, Y.; Lin, C.; Lee, J.J. Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interact. Learn. Environ.* **2017**, *6*, 778–791. [CrossRef]
6. Cai, S.; Liu, C.; Wang, T.; Liu, E.; Liang, J.C. Effects of learning physics using Augmented Reality on students' self-efficacy and conceptions of learning. *Br. J. Educ. Technol.* **2020**, *52*, 235–251. [CrossRef]
7. Bacca Acosta, J.L.; Baldiris Navarro, S.M.; Fabregat Gesa, R.; Kinshuk, K. Framework for designing motivational augmented reality applications in vocational education and training. *Australas. J. Educ. Technol.* **2019**, *35*, 102–117. [CrossRef]
8. Reid, N.; Shah, I. The role of laboratory work in university chemistry. *Chem. Educ. Res. Pract.* **2007**, *8*, 172–185. [CrossRef]
9. Wojciechowski, R.; Cellary, W. Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Comput. Educ.* **2013**, *68*, 570–585. [CrossRef]
10. Erbas, C.; Demirer, V. The effects of augmented reality on students' academic achievement and motivation in a biology course. *J. Comput. Assist. Learn.* **2019**, *35*, 450–458. [CrossRef]
11. Khan, T.; Johnston, K.; Ophoff, J. The Impact of an Augmented Reality Application on Learning Motivation of Students. *Adv. Hum. Comput. Interact.* **2019**, *2019*, 7208494. [CrossRef]
12. Huwer, J.; Barth, C.; Siol, A.; Eilks, I. Combining reflections on education for sustainability and digitalization—Learning with and about the sustainable use of tablets along an augmented reality learning environment. *Chemkon* **2021**, *28*, 235–240. [CrossRef]
13. Akçayır, M.; Akçayır, G.; Pektaş, H.M.; Ocak, M.A. Augmented reality in science laboratories: The effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. *Comput. Hum. Behav.* **2016**, *57*, 334–342. [CrossRef]
14. Sahin, D.; Yilmaz, R.M. The effect of Augmented Reality Technology on middle school students' achievements and attitudes towards science education. *Comput. Educ.* **2020**, *144*, 103710. [CrossRef]
15. Akçayır, M.; Akçayır, G. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educ. Res. Rev.* **2017**, *20*, 1–11. [CrossRef]
16. Albuquerque, G.; Sonntag, D.; Bodensiek, O.; Behlen, M.; Wendorff, N.; Magnor, M. A Framework for Data-Driven Augmented Reality. In *Augmented Reality, Virtual Reality, and Computer Graphics*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 71–83.
17. Cai, S.; Chiang, F.-K.; Wang, X. Using the Augmented Reality 3D Technique for a Convex Imaging Experiment in a Physics Course. *Int. J. Eng. Educ.* **2013**, *29*, 856–865.

18. Osman, S. The Effect of Augmented Reality Application toward Student's Learning Performance in PC Assembly. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 401–407. [CrossRef]

19. Salmi, H.; Thuneberg, H.; Vainikainen, M.-P. Making the invisible observable by Augmented Reality in informal science education context. *Int. J. Sci. Educ. Part B* **2016**, *7*, 253–268. [CrossRef]

20. Allcoat, D.; Hatchard, T.; Azmat, F.; Stansfield, K.; Watson, D.; von Mühlenen, A. Education in the Digital Age: Learning Experience in Virtual and Mixed Realities. *J. Educ. Comput. Res.* **2021**, *59*, 795–816. [CrossRef]

21. Buchner, J.; Kerres, M. Media comparison studies dominate comparative research on augmented reality in education. *Comput. Educ.* **2023**, *195*, 104711. [CrossRef]

22. Chao, J.; Chiu, J.L.; DeJaegher, C.J.; Pan, E.A. Sensor-Augmented Virtual Labs: Using Physical Interactions with Science Simulations to Promote Understanding of Gas Behavior. *J. Sci. Educ. Technol.* **2015**, *25*, 16–33. [CrossRef]

23. Chiu, J.L.; DeJaegher, C.J.; Chao, J. The effects of augmented virtual science laboratories on middle school students' understanding of gas properties. *Comput. Educ.* **2015**, *85*, 59–73. [CrossRef]

24. De Micheli, A.J.; Valentin, T.; Grillo, F.; Kapur, M.; Schuerle, S. Mixed Reality for an Enhanced Laboratory Course on Microfluidics. *J. Chem. Educ.* **2022**, *99*, 1272–1279. [CrossRef]

25. Fidan, M.; Tuncel, M. Integrating augmented reality into problem based learning: The effects on learning achievement and attitude in physics education. *Comput. Educ.* **2019**, *142*, 103635. [CrossRef]

26. Garzón, J.; Acevedo, J. Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educ. Res. Rev.* **2019**, *27*, 244–260. [CrossRef]

27. Hsiao, K.F.; Chen, N.S.; Huang, S.Y. Learning while exercising for science education in augmented reality among adolescents. *Interact. Learn. Environ.* **2012**, *20*, 331–349. [CrossRef]

28. Lu, S.-J.; Liu, Y.-C.; Chen, P.-J.; Hsieh, M.-R. Evaluation of AR embedded physical puzzle game on students' learning achievement and motivation on elementary natural science. *Interact. Learn. Environ.* **2018**, *28*, 451–463. [CrossRef]

29. Tarng, W.; Lin, Y.-J.; Ou, K.-L. A Virtual Experiment for Learning the Principle of Daniell Cell Based on Augmented Reality. *Appl. Sci.* **2021**, *11*, 762. [CrossRef]

30. Núñez-Redó, M.; Quirós, R.; Núñez, I.; Carda, J.; Camahort, E. Collaborative augmented reality for inorganic chemistry education. In Proceedings of the 5th WSEAS/IASME International Conference on Engineering Education, Vouliagmeni, Greece, 25–27 August 2008; pp. 271–277.

31. Radu, I.; Schneider, B.; Machinery, A.C. What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019.

32. Czok, V.; Krug, M.; Müller, S.; Huwer, J.; Kruse, S.; Müller, W.; Weitzel, H. A Framework for Analysis and Development of Augmented Reality Applications in Science and Engineering Teaching. *Educ. Sci.* **2023**, *13*, 926. [CrossRef]

33. Sirakaya, M.; Alsancak Sirakaya, D. Trends in Educational Augmented Reality Studies: A Systematic Review. *Malays. Online J. Educ. Technol.* **2018**, *6*, 60–74. [CrossRef]

34. Thees, M.; Kapp, S.; Strzys, M.P.; Beil, F.; Lukowicz, P.; Kuhn, J. Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Comput. Hum. Behav.* **2020**, *108*, 106316. [CrossRef]

35. Cai, S.; Wang, X.; Chiang, F.-K. A case study of Augmented Reality simulation system application in a chemistry course. *Comput. Hum. Behav.* **2014**, *37*, 31–40. [CrossRef]

36. Jones, L.L.; Kelly, R.M. Visualization: The Key to Understanding Chemistry Concepts. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; ACS Symposium Series; American Chemical Society: Washington, DC, USA, 2015; Volume 1208, pp. 121–140.

37. Rodríguez, F.C.; Frattini, G.; Krapp, L.F.; Martinez-Hung, H.; Moreno, D.M.; Roldán, M.; Salomón, J.; Stemkoski, L.; Traeger, S.; Dal Peraro, M.; et al. MoleculARweb: A Web Site for Chemistry and Structural Biology Education through Interactive Augmented Reality out of the Box in Commodity Devices. *J. Chem. Educ.* **2021**, *98*, 2243–2255. [CrossRef]

38. Domínguez Alfaro, J.L.; Gantois, S.; Blattgerste, J.; De Croon, R.; Verbert, K.; Pfeiffer, T.; Van Puyvelde, P. Mobile Augmented Reality Laboratory for Learning Acid–Base Titration. *J. Chem. Educ.* **2022**, *99*, 531–537. [CrossRef]

39. Fombona-Pascual, A.; Fombona, J.; Vicente, R. Augmented Reality, a Review of a Way to Represent and Manipulate 3D Chemical Structures. *J. Chem. Inf. Model.* **2022**, *62*, 1863–1872. [CrossRef]

40. Wong, C.H.S.; Tsang, K.C.K.; Chiu, W.K. Using Augmented Reality as a Powerful and Innovative Technology to Increase Enthusiasm and Enhance Student Learning in Higher Education Chemistry Courses. *J. Chem. Educ.* **2021**, *98*, 3476–3485. [CrossRef]

41. Mystakidis, S.; Fragkaki, M.; Filippousis, G. Ready Teacher One: Virtual and Augmented Reality Online Professional Development for K-12 School Teachers. *Computers* **2021**, *10*, 134. [CrossRef]

42. Teichrew, A.; Erb, R. How augmented reality enhances typical classroom experiments: Examples from mechanics, electricity and optics. *Phys. Educ.* **2020**, *55*, 065029. [CrossRef]

43. Wahyu, Y.; Suastra, I.W.; Sadia, I.W.; Suarni, N.K. The Effectiveness of Mobile Augmented Reality Assisted STEM-Based Learning on Scientific Literacy and Students' Achievement. *Int. J. Instr.* **2020**, *13*, 343–356. [CrossRef]

44. Eriksen, K.; Nielsen, B.E.; Pittelkow, M. Visualizing 3D Molecular Structures Using an Augmented Reality App. *J. Chem. Educ.* **2020**, *97*, 1487–1490. [CrossRef]

45. Pan, Z.; López, M.F.; Li, C.; Liu, M. Introducing augmented reality in early childhood literacy learning. *Res. Learn. Technol.* **2021**, *29*, 2539. [CrossRef]

46. Wu, H.K.; Lee, S.W.Y.; Chang, H.Y.; Liang, J.C. Current status, opportunities and challenges of augmented reality in education. *Comput. Educ.* **2013**, *62*, 41–49. [CrossRef]

47. Li, F.; Wang, X.; He, X.N.; Cheng, L.; Wang, Y.Y. How augmented reality affected academic achievement in K-12 education—A meta-analysis and thematic-analysis. *Interact. Learn. Environ.* **2021**, *31*, 5582–5600. [CrossRef]

48. Singh, G.; Mantri, A.; Sharma, O.; Dutta, R.; Kaur, R. Evaluating the impact of the augmented reality learning environment on electronics laboratory skills of engineering students. *Comput. Appl. Eng. Educ.* **2019**, *27*, 1361–1375. [CrossRef]

49. Lin, T.J.; Duh, H.B.L.; Li, N.; Wang, H.Y.; Tsai, C.C. An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. *Comput. Educ.* **2013**, *68*, 314–321. [CrossRef]

50. Penn, M.; Ramnarain, U. South African university students' attitudes towards chemistry learning in a virtually simulated learning environment. *Chem. Educ. Res. Pract.* **2019**, *20*, 699–709. [CrossRef]

51. Ling, Y.; Zhu, P.; Yu, J. Which types of learners are suitable for augmented reality? A fuzzy set analysis of learning outcomes configurations from the perspective of individual differences. *Educ. Technol. Res. Dev.* **2021**, *69*, 2985–3008. [CrossRef]

52. Sáez-López, J.M.; Cózar-Gutiérrez, R.; González-Calero, J.A.; Gómez Carrasco, C.J. Augmented Reality in Higher Education: An Evaluation Program in Initial Teacher Training. *Educ. Sci.* **2020**, *10*, 26. [CrossRef]

53. Wyss, C.; Furrer, F.; Degonda, A.; Bührer, W. Augmented Reality in der Hochschullehre. *Medien. Z. Theor. Prax. Medien.* **2022**, *47*, 118–137. [CrossRef]

54. Haas, B.; Lavicza, Z.; Houghton, T.; Kreis, Y. Can you create? Visualising and modelling real-world mathematics with technologies in STEAM educational settings. *Curr. Opin. Behav. Sci.* **2023**, *52*, 101297. [CrossRef]

55. Wang, M.; Callaghan, V.; Bernhardt, J.; White, K.; Peña-Rios, A. Augmented reality in education and training: Pedagogical approaches and illustrative case studies. *J. Ambient Intell. Humaniz. Comput.* **2017**, *9*, 1391–1402. [CrossRef]

56. Romano, M.; Díaz, P.; Aedo, I. Empowering teachers to create augmented reality experiences: The effects on the educational experience. *Interact. Learn. Environ.* **2023**, *31*, 1546–1563. [CrossRef]

57. Vogelsang, C.; Finger, A.; Laumann, D.; Thyssen, C. Vorerfahrungen, Einstellungen und motivationale Orientierungen als mögliche Einflussfaktoren auf den Einsatz digitaler Werkzeuge im naturwissenschaftlichen Unterricht. *Z. Didakt. Naturwiss.* **2019**, *25*, 115–129. [CrossRef]

58. Tesch, M.; Duit, R. Experimentieren im Physikunterricht—Ergebnisse einer Videostudie. *Z. Didakt. Naturwiss.* **2004**, *10*, 51–69.

59. Duit, R.; Tesch, M. On the role of the experiment in science teaching and learning—Visions and the reality of instructional practice. In *HSci 2010: 7th International Conference Hands-on Science "Bridging the Science and Society Gap", 25–31 July 2010, Rethimno, Greece*; Kalogiannakis, M., Stavrou, D., Michaelides, P.G., Eds.; The University of Crete: Rethymno, Greece, 2010.

60. Krug, M.; Czok, V.; Müller, S.; Weitzel, H.; Huwer, J.; Kruse, S.; Müller, W. AR in science education—An AR based teaching-learning scenario in the field of teacher education. *Chemkon* **2022**, *29*, 312–318. [CrossRef]

61. Sweller, J. Implications of Cognitive Load Theory for Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2005; Volume 27, pp. 27–42.

62. Mayer, R.E. *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2014. [CrossRef]

63. Henne, A.; Möhrke, P.; Thoms, L.-J.; Huwer, J. Implementing Digital Competencies in University Science Education Seminars Following the DiKoLAN Framework. *Educ. Sci.* **2022**, *12*, 356. [CrossRef]

64. Krug, M.; Thoms, L.-J.; Huwer, J. Augmented Reality in the Science Classroom—Implementing Pre-Service Teacher Training in the Competency Area of Simulation and Modeling According to the DiKoLAN Framework. *Educ. Sci.* **2023**, *13*, 1016. [CrossRef]

65. Becker, S.; Meßinger-Koppelt, J.; Thyssen, C.; Joachim Herz, S. Digitale Basiskompetenzen. In *Orientierungshilfe und Praxisbeispiele für die Universitäre Lehramtsausbildung in den Naturwissenschaften*; Auflage, Ed.; Joachim Herz Stiftung: Hamburg, Germany, 2020.

66. Kotzebue, L.v.; Meier, M.; Finger, A.; Kremser, E.; Huwer, J.; Thoms, L.-J.; Becker, S.; Bruckermann, T.; Thyssen, C. The Framework DiKoLAN (Digital Competencies for Teaching in Science Education) as Basis for the Self-Assessment Tool DiKoLAN-Grid. *Educ. Sci.* **2021**, *11*, 775. [CrossRef]

67. DiBaNa. Digital Basic Competencies for Science Teachers in an Online Platform for Acquiring Digital Teaching Competencies. Available online: https://www.dibana.de (accessed on 31 October 2023).

68. Autodesk, Inc. *Tinkercad*, 1.4; Autodesk, Inc.: San Francisco, CA, USA, 2022.

69. Zappar, Ltd. *Zapworks*, v6.5.34-stable; Zappar, Ltd.: London, UK, 2022.

70. International GeoGebra Institute. GeoGebra 3D; 5.0.744.A Web-App to Draw 3D Functions and Discover 3D Geometry with a 3D Graphing Calculator. Available online: https://www.geogebra.org/ (accessed on 31 October 2023).

71. De Backere, J.; Zambri, M. *LeARnCHEM*, 1.1. Available online: https://apps.apple.com/ca/app/learnchem/id1634480997(accessed on 31 October 2023).

72. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]

73. Fleiss, J.L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **1971**, *76*, 378–382. [CrossRef]

74. Krippendorff, K. Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educ. Psychol. Meas.* **2016**, *30*, 61–70. [CrossRef]

75. Sertdemir, Y.; Burgut, H.R.; Alparslan, Z.N.; Unal, I.; Gunasti, S. Comparing the methods of measuring multi-rater agreement on an ordinal rating scale: A simulation study with an application to real data. *J. Appl. Stat.* **2013**, *40*, 1506–1519. [CrossRef]

76. R Core Team. *R: A Language and Environment for Statistical Computing*; 4.3.1; R Foundation for Statistical Computing: Vienna, Austria, 2023.
77. Microsoft Corporation. *Microsoft Excel*, 16.78.3; Microsoft Corporation: Redmond, WA, USA, 2023.
78. Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika* **1945**, *10*, 255–282. [CrossRef]
79. Cronbach, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika* **1951**, *16*, 297–334. [CrossRef]
80. leARnCHEM. Available online: https://apps.apple.com/de/app/learnchem/id1634480997 (accessed on 31 October 2023).