Analyzing the Effective Use of Augmented Reality Glasses in University Physics Laboratory Courses for the Example Topic of Optical Polarization

Daniel Laumann[1,2] · Paul Schlummer[1] · Adrian Abazi[3] · Rasmus Borkamp[4] · Jonas Lauströer[4] ·
Wolfram Pernice[2,5] · Carsten Schuck[2,3] · Reinhard Schulz-Schaeffer[4] · Stefan Heusler[1]

Abstract

For nearly two decades, augmented reality (AR) has found diverse applications in education, particularly in science education, where its efficacy has been supported by relevant theories and many empirical studies. However, previous studies have revealed the following research deficit: While AR technology appears to influence learning-related variables, at the time of this study only few research on the use of AR glasses in physics, a discipline for which this technology seems particularly promising in the context of laboratory experiments, has been found. Thus, the present study uses an experimental comparison group design to investigate the question of how the use of AR glasses in a physics laboratory experiment (compared to in a learning setting without AR) influences students' motivation to learn, their cognitive load during the learning process and their learning achievement. The study (sample size $N = 75$) investigated the impact of AR glasses in a physics laboratory experiment on optical polarization. Results align with prior research, indicating heightened motivation among learners using AR applications. However, the absence of a significant difference in cognitive load between AR and non-AR learners was unexpected. Despite expectations based on spatial contiguity, learners with AR showed no advantage in learning achievement, challenging existing meta-analyses in physics education. These findings suggest a need to shift focus from surface features, like specific AR technology, to the content design of AR applications. Future studies should analyze the deep structure of AR applications, identifying features conducive to learning.

Keywords Augmented Reality · Physics Education · Lab Experiment

Introduction

Augmented reality (AR) as a technology gained broader recognition in the mid-1990s and is described as being part of "mixed reality" technologies on the reality-virtuality continuum (Milgram & Kishino, 1994). Here, "an AR system supplements the real world with virtual objects that appear to coexist in the same space as the real world" (Azuma et al., 2001, p. 34). Further characteristics of AR include the three-dimensional (3D) alignment of real and virtual objects as well as real-time interactivity of the representations (Azuma, 1997).

While early central publications about AR focused on application areas such as medicine or entertainment (Azuma et al., 2001; Milgram & Kishino, 1994), later publications offered the perspective of using AR for education (Szalavári et al., 1998).

AR in Education

Today, AR has been widely applied to teaching and learning, with most applications in the natural sciences followed by social sciences (Chang et al., 2022). AR can be implemented using different technologies and various devices (mobile devices such as smartphones or tablet computers, AR glasses) or even without any personal equipment (spatial AR)

✉ Daniel Laumann
daniel.laumann@uni-muenster.de

[1] Institute of Physics Education, University of Münster, Wilhelm-Klemm-Straße 10, 48149 Münster, Germany

[2] Institute of Physics, University of Münster, Wilhelm-Klemm-Straße 10, 48149 Münster, Germany

[3] Department for Quantum Technology, University of Münster, Heisenbergstraße 11, 48149 Münster, Germany

[4] Department Design, HAW Hamburg, Finkenau 35, 22081 Hamburg, Germany

[5] Kurchhoff Institute for Physics, University of Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany

(Buchner et al., 2022). Currently, the most common applications are those in which AR content is accessed via mobile devices (Xu et al., 2022; Zhang et al., 2022). Although several applications for AR glasses exist, hardly any empirical findings have been reported on this technology.

Affective Characteristics

Of the findings that have been reported, those for educational contexts without differentiation by discipline are related to affective characteristics (Bacca et al., 2014; Cao & Yu, 2023; Chang et al., 2022), such as attitudes or motivation, when AR-based learning environments are compared to traditional (non-AR) learning environments. Concerning these affective variables, results have shown that learning in AR-based learning environments is perceived more positively than traditional non-AR learning environments (Hedge's $g = 0.49$, $p < .001$; Chang et al., 2022) with small effect sizes.[1] A more differentiated analysis of affective variables has indicated that attitudes toward AR-based learning environments are more positive than those for non-AR learning environments (Cohen's $d = 1.08$, $p = .001$; Cao & Yu, 2023), but no differences have been identified regarding learning motivation (Cohen's $d = 0.25$, $p = .120$; Cao & Yu, 2023). For example, studies in the field of language learning that examined the duration of an intervention have suggested that as the duration of the intervention increases, AR may have a stronger positive influence on motivation (Cai et al., 2022).

Cognitive Load

Different studies have analyzed learners' cognitive load (Sweller, 1988, 2011), which is an important predictor of learning achievement (Sweller et al., 2011). Cognitive load is directly related to the Cognitive Theory of Multimedia Learning (Mayer, 2009), which provides insights into the design of multimedia learning environments. In a study examining students' cognitive load for AR-based learning environments, cognitive load was reported by learners as being lower than or equal to the cognitive load in non-AR learning environments, and higher learning achievement was also shown for the AR-based learning environment (Buchner et al., 2022). This has also been confirmed by other comparative analyses with less differentiated consideration and joint analysis of learning outcomes including both cognitive load and learning achievement (Zhang et al., 2022).

---

[1] The specification of effect sizes in the manuscript is based on Cohen (1988). Therefore, the following interpretive ranges for effect size apply to all results. Cohen's d: $|d| \geq 0.20$ small effect, $|d| \geq 0.50$ medium effect, $|d| \geq 0.80$ large effect. Hedge's g: $|g| \geq 0.20$ small effect, $|g| \geq 0.50$ medium effect, $|g| \geq 0.80$ large effect.

Learning Achievement

Numerous overview studies (Bacca et al., 2014; Cao & Yu, 2023; Chang et al., 2022; Zhang et al., 2022) have demonstrated advantages of AR-based learning environments for learning achievement with medium effect sizes (Hedge's $g = 0.64$, $p < .001$) according to a recent second-order meta-analysis (Malone et al., 2023) comparing AR-based learning environments vs. traditional learning environments without AR. The mean effect size can be interpreted in such a way that all individual analyses demonstrate medium to large effects and are therefore consistent regarding the positive impact of AR on learning achievement (Hedge's $g_{min} = 0.42$, $p = .04$; Hedge's $g_{max} = 1.69$, $p = .03$).

AR may positively influence learning according to educational sciences due to the coexistence of the real world and virtual objects causing an impression of immersion for learners. Here, immersion is described as a construct related to flow experience (Csikszentmihalyi, 1975, 1990). According to different studies (Antonietti et al., 2000; Conole & Dyke, 2004; Dalgarno & Lee, 2010; Dunleavy et al., 2009), immersion via AR involves corresponding affective variables that seem to be positively related to learning outcomes.

Further, cognitive psychological reasoning according to the Cognitive Theory of Multimedia Learning (Mayer, 2009) shows that integrating different multiple representations with high spatial and temporal contiguity is particularly effective for learning (Ginns, 2006; Schroeder & Cenkci, 2018). For interactive representations, high contiguity can be realized easily with AR.

AR in Science Education

As indicated above, theories (Cognitive Load Theory, Cognitive Theory of Multimedia Learning) and empirical findings suggest that AR has strong potential for use in learning environments. AR appears particularly promising for science education, where field studies and virtual information can be enriched by adding visualizations of associated models when conducting experiments. This rationale has been used in numerous recent studies empirically analyzing AR when applied in science education, such as in physics, particularly for conducting experiments in lab settings (Strzys et al., 2018; Kapp et al., 2019; Schlummer et al., 2023; Wagner et al., 2023).

Affective Characteristics

Review studies in science education have generally found evidence for a positive effect of AR-based learning environments on affective variables, such as motivation or attitudes, compared to non-AR learning environments (Ibáñez & Delgado-Kloos, 2018; Sırakaya & Sırakaya, 2020). Individual studies have shown that no differences

for positive emotions between learning environments occur with and without AR (Elford et al., 2022). Also, other studies on flow experience have shown advantages of AR-based learning environments (Ibáñez et al., 2014). Overall, however, structured data on the affective outcomes of AR-based learning environments compared to non-AR learning environments are scarce (Xu et al., 2022).

Cognitive Load

Similar findings as for affective traits have also emerged for analyses of cognitive load. Again, review articles have suggested that AR seems to be able to reduce cognitive load (Ibáñez & Delgado-Kloos, 2018; Sırakaya & Sırakaya, 2020). However, these reviews contain neither concrete statements on the reduction of cognitive load regarding different types of cognitive load (extraneous, germane, intrinsic) nor quantitative information on effect sizes. Examining individual studies suggests, analogous to studies of cognitive load without differentiation by discipline for the field of science education, that no difference in cognitive load exists between AR-based learning environments and non-AR learning environments (Altmeyer et al., 2020; Elford et al., 2022) or that extraneous cognitive load is only somewhat reduced by AR-based learning environments (Thees et al., 2020).

Learning Achievement

In science education, studies have indicated that students in AR-based learning environments show higher learning achievement regarding their ability to remember information, their conceptual understanding and their effectiveness in problem-solving procedures compared to non-AR learning environments (Ibáñez & Delgado-Kloos, 2018). This is also supported by a second-order meta-analysis (Malone et al., 2023) showing that AR learning environments positively influence learning with medium effect sizes (Hedge's $g = 0.66$, $p < .001$). However, differences have emerged for different disciplines within science education (Xu et al., 2022). For example, no advantages have been found in biology for AR-based learning environments, whereas advantages in physics have medium effect sizes (Hedge's $g = 0.64$) and advantages in the earth sciences are the strongest (Hedge's $g = 1.45$). One suggested reason for these strong effects in earth sciences, which includes astronomy and geography, is that spatial skills are crucial and can be promoted particularly well by AR, while in biology, methods with real observations are important and cannot be taught better via AR (Xu et al., 2022).

Furthermore, while studies on AR based on mobile devices dominate in the field of science education and across all disciplines, at the time of this study only one published article was found that analyzed the effects of AR glasses in learning settings in the field of science education (Thees et al., 2020). This study found no differences in the learning effectiveness of AR compared to traditional settings, which seems to be highly important for science education and especially for physics. In physics, using AR to promote interaction with physics concepts (Azuma et al., 2001) can be effectively achieved with AR glasses when AR is used as an additive element in experimentation, since students' hands remain free for interacting with the experiment. In addition, numerous physics applications already make use of AR glasses (Strzys et al., 2018; Kapp et al., 2019; Schlummer et al., 2023; Wagner et al., 2023).

Research Questions

The previous discussion concerning the state of science education research on the use of AR in learning processes reveals that only a few systematic findings are available regarding combined studies of students' affective responses to AR-based learning environments and their cognitive load, although both aspects are significant for learning success. Furthermore, previous studies have focused on AR used on mobile devices (smartphones or tablet computers). However, the use of AR glasses in laboratory settings appears to be promising based on findings from educational psychology, since only this technology allows for unrestricted practical action and manual interaction as well as a high degree of spatial contiguity between experimental setup and model-related visualizations. Nevertheless AR glasses have hardly been considered for empirical studies. To provide additional findings for the effects of AR glasses on affective characteristics, cognitive load and learning achievement in science education and to include AR glasses as an application, the following research questions will be investigated exemplarily for a physics topic, namely optical polarization:

How does the use of AR glasses in a laboratory experiment for optical polarization compared to a learning setting without AR influence students'…

RQ1… motivation to learn?
RQ2… cognitive load during the learning process?
RQ3… learning achievement?

Methods

The experimental study (Fig. 1) compared learning for an AR group (laboratory experiment with additional use of AR) and a non-AR group (laboratory experiment with traditional supplemental material, namely further traditional experimental equipment and printed teaching material) with equal information and representations, see "Laboratory
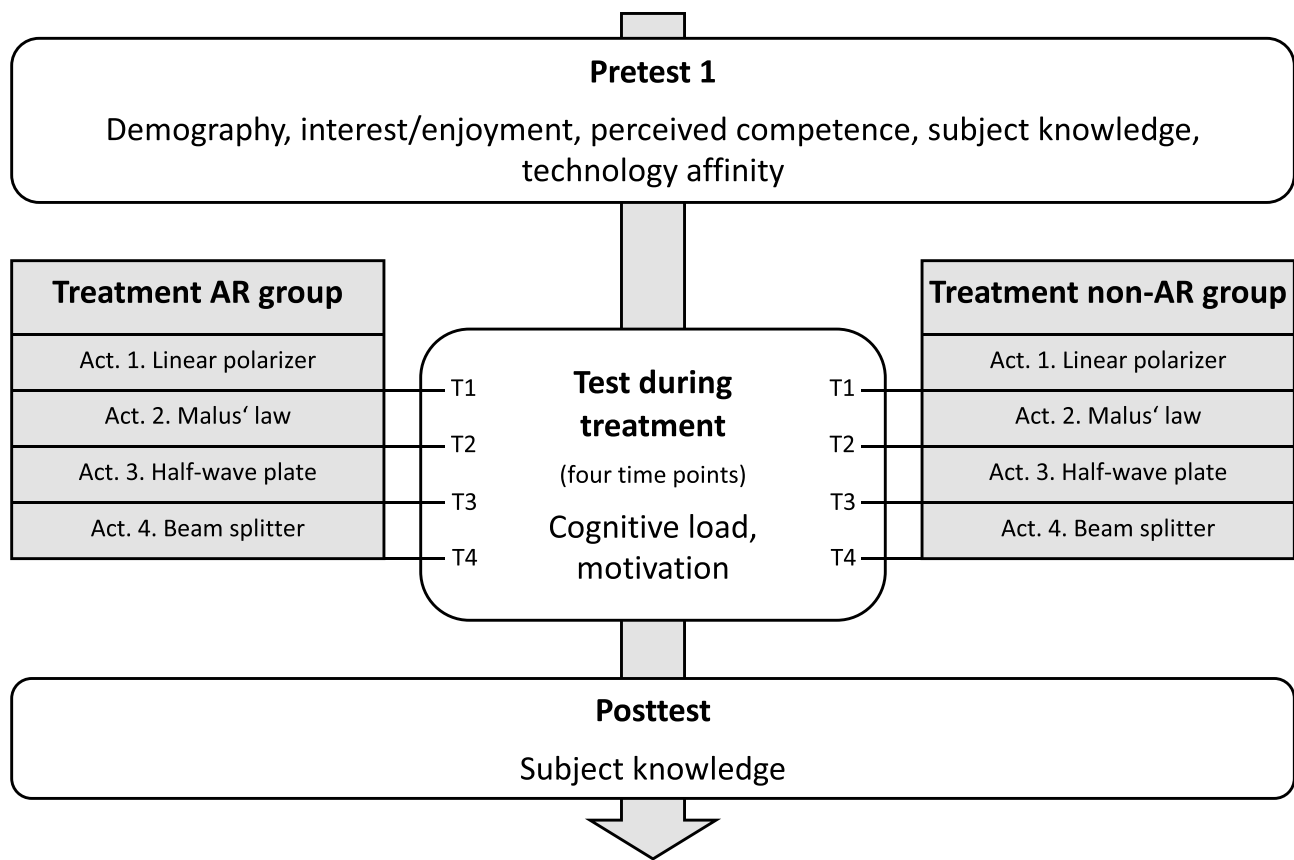
Fig. 1 Flowchart of experimental comparison group design of the study comparing learning in a physics laboratory course for a laboratory experiment supplemented by augmented reality content (AR group) and a laboratory experiment with traditional supplemental material (non-AR group)

experiment: Optical polarization" section. The treatment comprised a 3-h laboratory experiment on optical polarization, where learners worked individually on individual experimental setups. The students were randomly assigned to the experimental groups.

Laboratory Experiment: Optical polarization

The laboratory experiment provided learners with a setup to investigate the properties of linearly polarized light and the functionalities of various optical components that are frequently used in optics laboratories (Schlummer et al., 2023). The main setup consisted of a diode laser whose intensity can be measured by means of a photodiode (Fig. 2).

Between the laser and the photodiode, up to three optical components could be installed. In a sequence of four different activities, the goal of the laboratory experiment was to:

Activity 1.Determine the degree of polarization of the used laser diode, as well as its angle of maximum emission.

Activity 2.Malus' law: Characterize the selective absorption properties of a linear polarizer by finding an appropriate functional dependency.
Activity 3.Investigate the effects of a half-wave plate on linearly polarized light and compare it to the behavior of a polarizer.
Activity 4.Characterize the two output channels of a polarizing beam splitter and explore how to use it as a measurement device in combination with a half-wave plate.

In each of the four activities, the learners would need to perform the necessary steps to create a functional configuration by installing the respective components, investigating the effects of rotating components within the setup qualitatively and taking appropriate data to describe the functionality of each component by means of appropriate parameters.

For the AR group, the learners had access to interactive model visualizations during their work with the experiment. These visualizations included written instructions on the current task as well as vector-representations for each

component in the setup and a real-time plot of measurement data, Fig. 2 top. Figure 3 shows the most relevant types of model representations in more detail.

In the specific example, partially polarized light from a laser diode coming from the left successively passes an initial polarizer (A), a half-wave-plate (B) and a second polarizer (C). The respective vector representations match with the rotation angles of the components set in the physical setup. Each diagram shows the polarization direction and relative intensity of the incoming light in dark green and for the transmitted light in light green.

While the general layout of the diagrams is the same for all components, there are several small differences between them indicating the different physical properties of the respective components. Before the light enters the initial polarizer (A) in Fig. 3, it is only partially polarized. Consequently, there is no well-defined vector for the incoming light in the first diagram. Instead, the polarization characteristics of the partially polarized laser diode is added as a dark green shape. This helps the user to set the initial polarizer to an angle that provides optimal intensity for the light beam. After passing the initial polarizer, the light can be assumed to be linearly polarized, as is indicated by the radiant green double arrow in the diagram.

Diagram (B) illustrates that the half wave plate rotates the polarization of the incoming light but does not change its intensity, as the dark and the light green double arrows have the same length. This is different in the third diagram, where it is clearly visible that the polarizer changes both the direction and the intensity of the transmitted light.

Quantitative data on the transmitted intensity can be visualized in the measurement diagram (D). In this case, the light intensity behind the setup is measured as a function of the half-wave plate's rotation angle.

The non-AR group received a workbook with written instructions which also contained static versions of the interactive vector representations available for the AR group. The workbook also provided a diagram template to be filled out for each task so that learners could create their own visualizations of the measured data (Fig. 2, bottom). This was done to ensure a fair comparison with respect to the live-diagram functionality of the AR group.

Setting and Participants

The present study considers data collected between April and July 2022 during a regular laboratory course for physics students (Bachelor of Science, Bachelor of Education). $N = 75$ students participated in the study. Students were randomly assigned to a treatment condition (AR group or non-AR group). Further information describing the sample can be found in Table 1. $N_{AR} = 19$ and $N_{Non-AR} = 20$ students indicated previous experience with immersive headsets, mainly in the context of VR gaming applications.

The learners in the AR group used a *Microsoft HoloLens 2* (mixed reality head-mounted display). The glasses are available for educational institutions such as universities and schools for a price of around USD 3,500. The glasses have a diagonal field of view of 52 degrees and provide interaction by hand gestures and eye-tracking with no need for handheld controllers (Fig. 4).

Data Collection and Analysis

Data was collected using an online questionnaire for the pretest and posttest. Due to the simplified handling during the experiment, respective data was collected by means of a paper–pencil questionnaire.

Motivation

Learners' motivation was measured using the Intrinsic Motivation Inventory (IMI; McAuley et al., 1989) in the German-language abridged version (Wilde et al., 2009) Specifically, the IMI subscales *interest/enjoyment* (IE; 3 items, 5-point Likert scale) and *perceived competence* (PC; 3 items, 5-point Likert scale) were used. We considered that the IMI explicitly covers the construct *motivation* very broadly and that the scales can also be used individually depending on the content fit (McAuley et al., 1989). Motivation was assessed during the treatment (four measurement time points; each after the four specific activities of the experiment described in Sect. 3.1.). Mean scores were taken for each subscale, and then total motivation was calculated as the sum of the mean scores for interest/enjoyment and perceived competence.

The analysis of the internal consistency shows very good to excellent values for both subscales by determining Cronbach's α ($\alpha_{IE} = .90$; $\alpha_{PC} = .89$; Cortina, 1993).

Cognitive Load

Cognitive load was measured as a three-dimensional construct divided into intrinsic load (IL), extraneous load (EL) and germane load (GL). This separate measurement of the individual contributions seems necessary because, based on theories such as the Cognitive Theory of Multimedia Learning (Mayer, 2009), it can be assumed that the design of the learning environment only influences the extraneous load part of the cognitive load (Mayer & Fiorella, 2014). In contrast to extraneous load, measurement of intrinsic load and germane load can be used to ensure that the learning environment does not differ in features other than technological support provided by AR and non-AR, respectively. A self-assessment instrument (7 items, 7-point Likert scale; Klepsch et al., 2017) was used to measure extraneous load with 3 items and intrinsic load and germane load with 2 items each. The dependent variables IL, EL and GL were
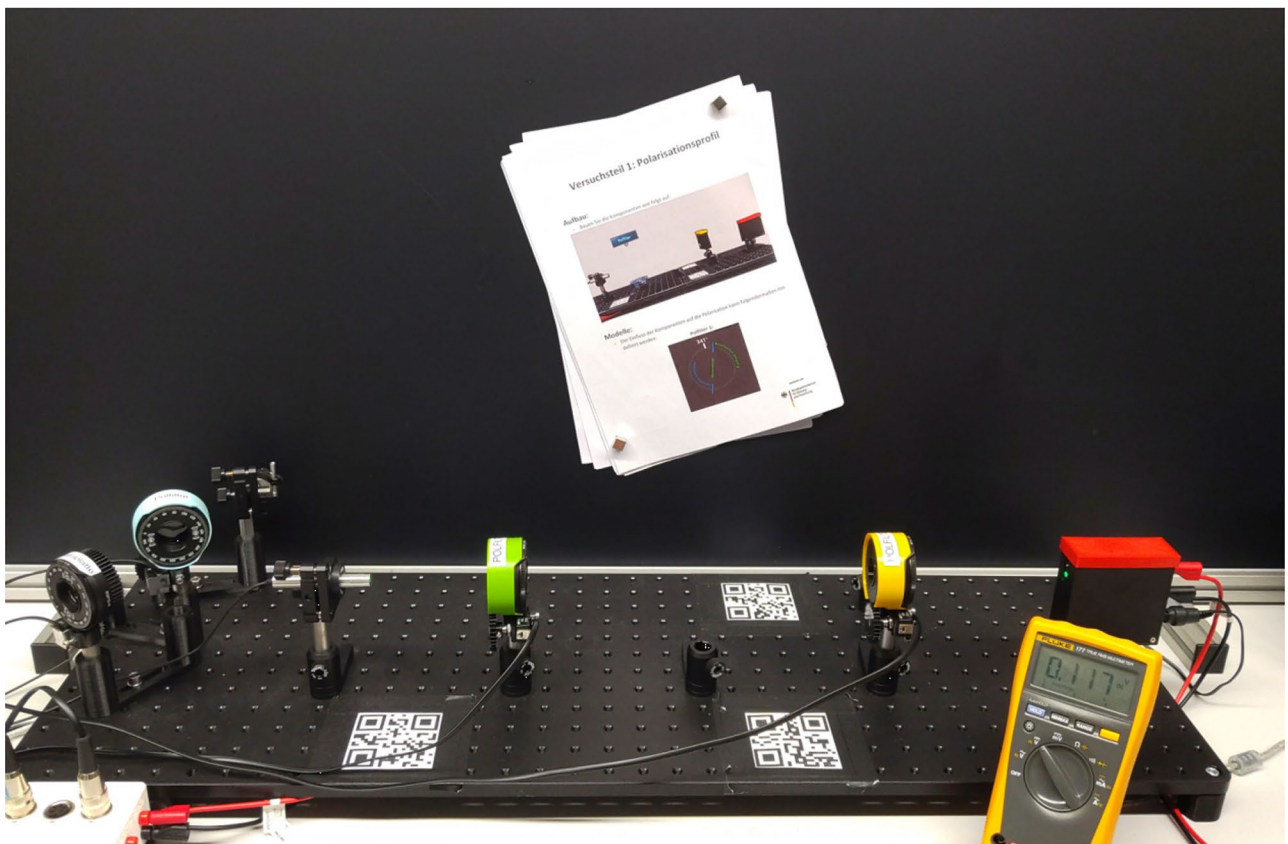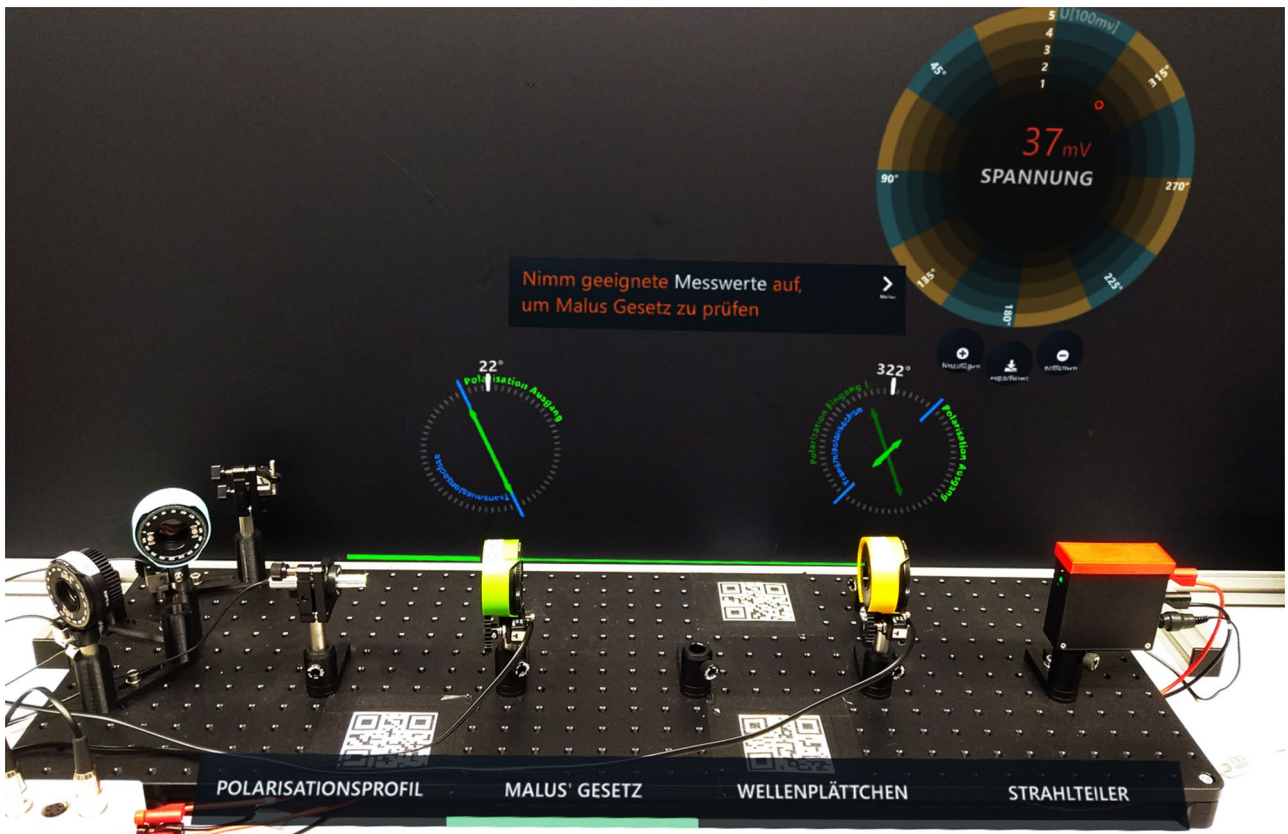
◀**Fig. 2** Experimental setup to investigate optical polarization for a laboratory experiment supplemented by augmented reality content (AR group; top) and a laboratory experiment with traditional supplemental material (non-AR group; bottom)

also measured after each of the four specific activities of the experiment described in "Laboratory experiment: Optical polarization" section. Mean scores were calculated for each subscale. Total cognitive load was calculated as the sum score of all three contributions.

Analysis of the internal consistency shows good to very good values for all three subscales (IL, EL and GL) by Cronbach's α (α_{IL} = .88; α_{EL} = .82; α_{GL} = .70; Cortina, 1993).

Learning Achievement

Learning achievement was measured by a self-developed test instrument on the learners' subject knowledge as a dependent variable for both groups. The use of a self-developed test instrument was intended to ensure that the specific learning gains of the experiment could be determined with a high fit to the content taught. The test included 19 multiple-choice items. While the items were used in their entirety to measure subject knowledge in the posttest, the measurement in the pretest was only for 8 items, because certain items were too specific to be answered by the learners before the laboratory course, such as the item shown in Fig. 5. The fit of the items to the content of the laboratory experiment was tested in previous studies in the same teaching–learning context. The items tested whether learners were able to recall the operation of the experimental components (retention) and transfer what they learned to other experimental configurations (transfer). The indication of total values was done by sum scores.

The pretest and posttest subscales had varying internal consistencies as described by Cronbach's α (α_{Pre} = 0.46; α_{Post} = 0.79; Cortina, 1993). The low consistency of the
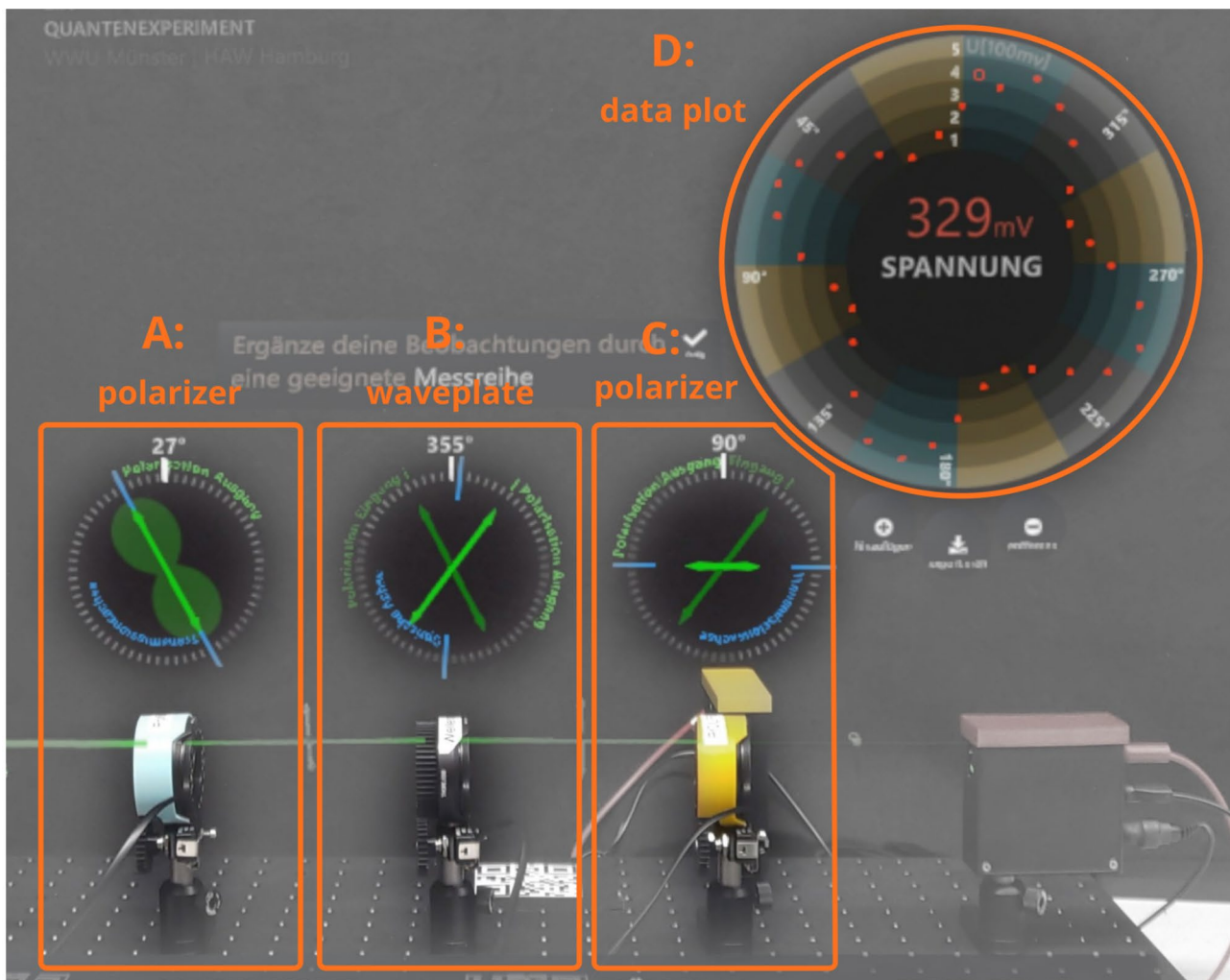
Fig. 3 Detailed view of the different forms of representations provided to the user. Fields A-C show the vector diagrams for the respective components underneath, Field D shows an example for a measurement diagram in this specific setup

Table 1 Sample description of the study and description of the study groups indicating students' semester of study as well as gender. (Only n = 37 students in the AR group were included in the calculation of mean values and standard deviations due to two participant's incorrect indication of semester)

Group	Gender	Semester of study		
		N	M	SD
AR	Total	39	4.11	0.70
	Female	12	3.82	0.60
	Male	27	4.23	0.71
Non-AR	Total	36	4.06	0.33
	Female	5	4.00	0.00
	Male	31	4.06	0.36

corresponding test on subject knowledge arose because different content was covered pretest, and these concepts were made clear only via the laboratory course itself; so, high values for Cronbach's α on the pretest were not necessarily expected (Taber, 2018; Stadler et al., 2021).

Further Variables

Demographic data used to describe the sample included gender, semester of study and prior experience using immersive headsets for VR and AR applications.

Given the analysis of the instructional impact of a new technology (AR), it seems useful to analyze learners' technology affinity (TA). There is evidence that positive technology affinity is positively correlated with learning in technology-based learning environments but negatively correlated with learning in traditional learning settings (Backhaus et al., 2019). Furthermore, previous studies also suggest that learners' attitudes towards education, including their motivation (Mills et al., 2013) as well as their learning intentions (Jin & Divitini, 2020), may be influenced by their technology affinity. Furthermore, there is no clear indication of the extent to which technology affinity correlates with learners' cognitive load, as some studies have found a corresponding correlation (Albus et al., 2021), while others did not (Binder et al., 2021;

Fig. 4 Photo of the learners working with the HoloLens 2 in the laboratory experiment supplemented by the visualizations of the AR

Which of the following electric field vectors fits
to the measured intensity distribution?
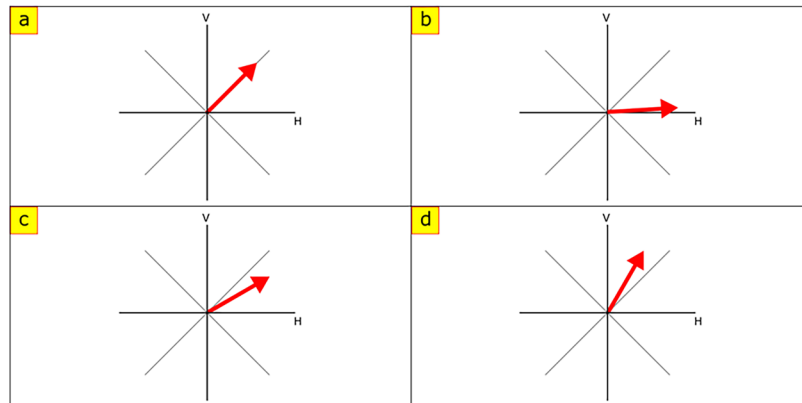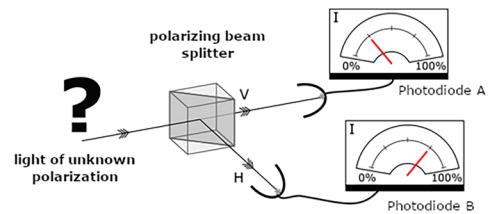
Fig. 5 Example item for the extended set of items used in the posttest. The participant was presented with a graphical representation of an experimental setup and asked to identify the direction of the polariza-tion vector that fits the given experimental results (the correct answer is option [c] in this case)

Laun et al., 2022). Therefore, learners' technology affinity was included as a potential covariate for the analyses of motivation, cognitive load and learning achievement. Technology affinity was analyzed using the two subscales *enthusiasm for technology* (ET; 5 items, 5-point Likert scale) and *(self-assessed) competence in using technology* (CT; 4 items, 5-point Likert scale) of the widely used TA-EG (Karrer et al., 2009; Siebert et al., 2022). The analysis of the internal consistency shows good values for the two subscales of technology affinity via Cronbach's α ($\alpha_{ET} = .84$; $\alpha_{CT} = .86$; Cortina, 1993).

Findings

The results for research questions 1–3 are described separately below. Data was analyzed with R (Version 4.2.3) in RStudio (Version 2022.07.1). Effect sizes were specified according to Cohen (1988): Wilcoxon effect size $|r| \geq .10$ small, $|r| \geq .30$ medium, $|r| \geq .50$ large effect size; partial $\eta^2 \geq .01$ small, $\eta^2 \geq .06$ medium, $\eta^2 \geq .14$ large effect size.

Before applying the parametric test procedures used below, the appropriateness of the assumption of a normal distribution was checked. The test was carried out for all

eight characteristics included as dependent variables (interest/enjoyment, perceived competence, total motivation, extraneous cognitive load, germane cognitive load, intrinsic cognitive load, total cognitive load, subject knowledge posttest) and divided into the two groups (AR group, non-AR group). Normality was assumed as determined by the Shapiro–Wilk test and upon visual inspection as of the quantile–quantile-plots.

AR vs. Non-AR Regarding Motivation (RQ1)

Following previous studies, learners' motivation in completing the laboratory course may be influenced by their technology affinity, as hypothesized in Sect. 3.3.4. To determine whether the learners' technology affinity should be considered in analyses of differences in motivation between the AR and non-AR groups, two forced-entry linear regressions, see Table 2, were carried out (independent variables: enthusiasm for technology, self-assessed competence in using technology; dependent variables: interest/enjoyment, perceived competence). Results showed that neither enthusiasm for technology nor competence in using technology significantly determined interest/enjoyment or perceived competence for the laboratory experiment. Therefore, analyses of group differences between AR and non-AR group were performed without any covariates.

Table 2 Fit parameters obtained from regression models for interest/enjoyment and perceived competence as dependent variables

Variable	$\beta(SD)$	t	p
Interest/enjoyment			
(Intercept)	-	6.03	$<.001$
Enthusiasm for technology	0.00(0.28)	0.01	.993
Competence in using technology	0.14(0.30)	0.96	.343
Adjusted R^2			-0.006
Perceived competence			
(Intercept)	-	7.40	$<.001$
Enthusiasm for technology	-0.02(0.24)	-0.16	.875
Competence in using technology	0.26(0.26)	1.80	.076
Adjusted R^2			0.037

The analysis of group differences (between-subjects factor) and time courses (within-subjects factor) regarding motivational characteristics was performed using type II mixed ANOVAs for interest/enjoyment, perceived competence, and total motivation as dependent variables; see Table 3. Results included the overall group differences between the AR group and non-AR group (group effect), the temporal development of motivational characteristics over the four measurement time points without considering the groups (time effect), and the corresponding interaction effect.

The analysis for group effects revealed that the students in the AR group rated their interest/enjoyment higher ($M = 3.69$, $SD = 0.65$) than students in the non-AR group ($M = 3.01$, $SD = 0.71$), and the mixed ANOVA showed that this is a large significant effect ($F(1.00, 70.00) = 10.90$, $p = .002$, partial $\eta^2 = 0.14$). Although the perceived competence of the AR group ($M = 3.57$, $SD = 0.88$) and the non-AR group ($M = 3.57$, $SD = 0.71$) did not differ significantly ($F(1.00, 70.00) = 0.01$, $p = .941$), there was a significant difference, with a small to medium effect size ($F(1.00, 70.00) = 4.40$, $p = .039$, partial $\eta^2 = .059$), in total motivation of the learners in the AR group ($M = 3.62$, $SD = 0.64$) compared to the non-AR group ($M = 3.29$, $SD = 0.70$). Since all motivational variables were assessed as mean scores of 3 items on 5-point Likert scales, results indicate that learners expressed medium to rather high ratings (Fig. 6). In addition, no significant time effect or interaction effect is present in the data.

AR vs. Non-AR Regarding Cognitive Load (RQ2)

The amount of cognitive load students experienced during the experiment might be related to learners' technology affinity, as explained in "Futher Variables" section, although this does not seem clear. To determine whether technology affinity should be considered in analyses of differences in cognitive load between the AR and non-AR groups, a forced-entry linear regression was carried out for each cognitive load component (independent variables: enthusiasm for technology, self-assessed competence in using technology; dependent variables: cognitive load components).

The linear regression, see Table 4, revealed that for the present data, neither enthusiasm for technology nor competence in using technology significantly determined intrinsic cognitive load, extraneous cognitive load or germane cognitive load for the laboratory experiment. Therefore, analyses of group differences between AR and non-AR groups were performed without taking any covariates into account.

The analysis of group differences and trajectories regarding cognitive load was performed using type II mixed ANOVAs for intrinsic cognitive load, extraneous cognitive load, germane cognitive load and total cognitive load as dependent variables, see Table 5. Results include overall group differences between the AR group and non-AR group (group effect), the temporal development of cognitive load over the four measurement time points without considering the groups (time effect), and the corresponding interaction effect.

The analysis for group effects revealed that the students in the AR group rated their intrinsic cognitive load as higher ($M = 3.12$, $SD = 1.04$) than students in the non-AR group ($M = 2.65$, $SD = 0.97$), and the mixed ANOVA showed that this is a significant effect of moderate size ($F(1.00, 70.00) = 4.28$, $p = .042$, partial $\eta^2 = 0.06$). The differences for extraneous

Table 3 Results for main effects and interaction effects (group differences) between AR group and non-AR group for time course of interest/enjoyment, perceived competence and total motivation (type II mixed ANOVAs)

Variable	Effect	F	$df1$	$df2$	p	Partial η^2
Interest/enjoyment	Group	10.90	1.00	70.00	.002	0.14
	Time	1.31	2.72	190.00	.275	-
	Interaction	0.51	2.72	190.00	.512	-
Perceived competence	Group	0.01	1.00	70.00	.941	-
	Time	0.95	3.00	210.00	.418	-
	Interaction	0.86	3.00	210.00	.465	-
Total motivation	Group	4.40	1.00	70.00	.039	0.06
	Time	1.33	2.67	186.91	.266	-
	Interaction	0.66	2.67	186.91	.562	-
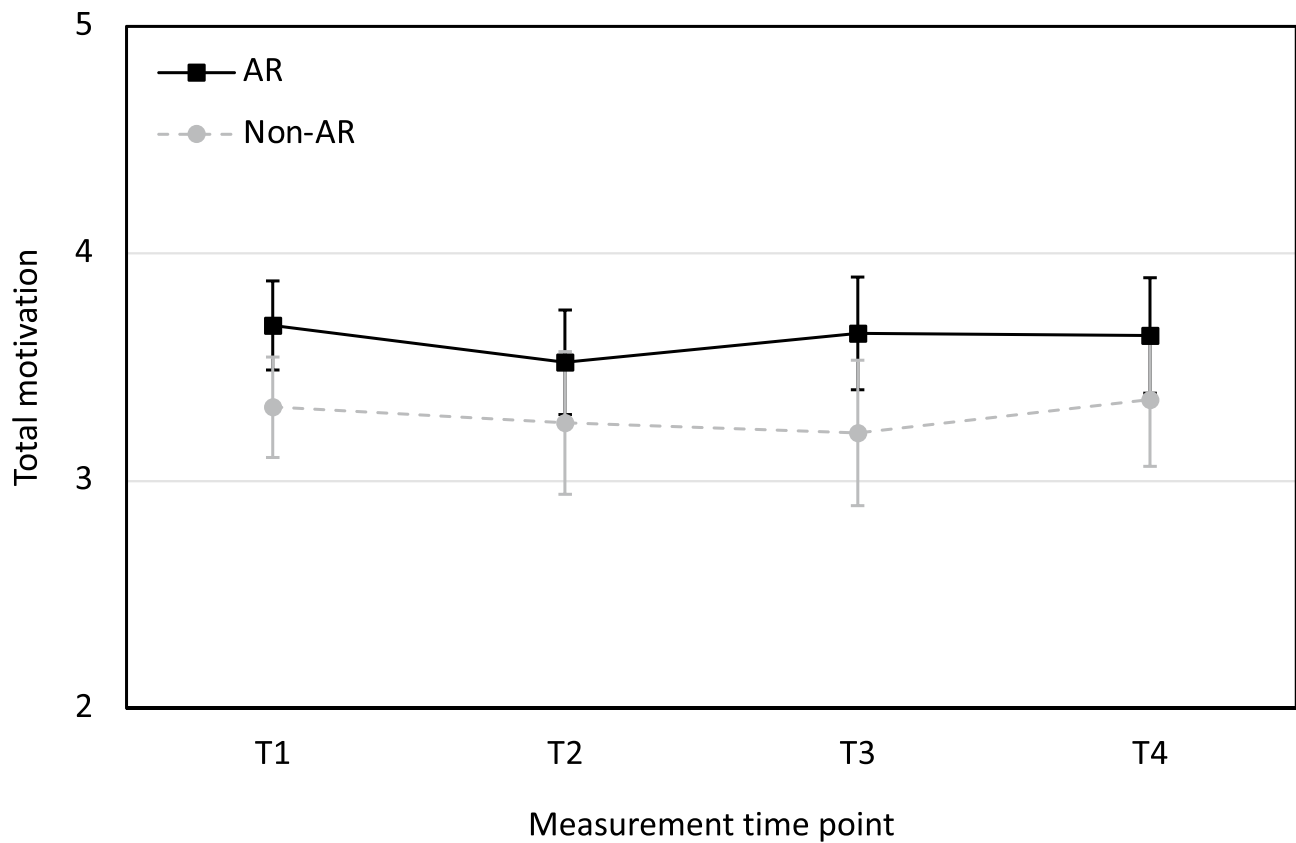
Fig. 6 Time course of total motivation over the four measurement time points for the AR and the non-AR groups

cognitive load between the AR group ($M = 3.19$, $SD = 1.13$) and the non-AR group ($M = 3.17$, $SD = 0.88$) as well as the differences for germane cognitive load between the AR group ($M = 4.73$, $SD = 1.01$) and the non-AR group ($M = 4.55$,

Table 4 Fit parameters obtained from regression models for intrinsic cognitive load, extraneous cognitive load and germane cognitive load as dependent variables

Variable	$\beta(SE)$	t	p
Intrinsic cognitive load			
(Intercept)	-	6.25	< .001
Enthusiasm for technology	0.13(0.32)	0.87	.387
Competence in using technology	-0.18(0.34)	-1.17	.247
Adjusted R^2			-0.008
Extraneous cognitive load			
(Intercept)	-	7.34	< .001
Enthusiasm for technology	0.03(0.31)	0.21	.836
Competence in using technology	-0.14(0.33)	-0.927	.357
Adjusted R^2			-0.012
Germane cognitive load			
(Intercept)	-	7.96	< 0.001
Enthusiasm for technology	0.18(0.36)	1.18	.240
Competence in using technology	-0.14(0.34)	-0.93	.357
Adjusted R^2			-0.007

$SD = 1.35$) are not significant (extraneous cognitive load: $F(1.00, 70.00) = 0.01$, $p = .910$; germane cognitive load: $F(1.00, 70.00) = 0.37$, $p = .545$). Also, total cognitive load differences between the AR group ($M = 3.68$, $SD = 0.82$ and the non-AR group ($M = 3.46$, $SD = 0.79$) are not significant ($F(1.00, 70.00) = 1.45$, $p = .233$).

Regarding time effects, results showed that intrinsic cognitive load ($F(2.64, 184.47) = 20.75$, $p < .001$, partial $\eta^2 = 0.23$), extraneous cognitive load ($F(2.58, 180.40) = 9.12$, $p < .001$, partial $\eta^2 = 0.12$) and total cognitive load ($F(2.62, 183.20) = 19.16$, $p < .001$, partial $\eta^2 = 0.22$) changed over the course of the laboratory experiment with moderate to large effect size (Fig. 7). Concerning interaction effects, we found no differences in the temporal development of cognitive load over the four measurement time points between the two groups.

AR vs. Non-AR Regarding Learning Achievement (RQ3)

We also analyzed possible effects on the students' learning achievement. Due to possible correlations between motivation and technology affinity to learning achievement, first a forced-entry preliminary linear regression was carried out with a model function that included subject knowledge in the posttest as an independent variable and subject knowledge in the pretest as a further dependent variable.

Table 5 Results for main effects and interaction effects (group differences) between AR group and non-AR group for time course of intrinsic cognitive load, extraneous cognitive load, germane cognitive load and total cognitive load (mixed ANOVAs)

Variable	Effect	F	$df1$	$df2$	p	Partial η^2
Intrinsic cognitive load	Group	4.28	1.00	70.00	.042	0.06
	Time	20.75	2.64	184.47	< .001	0.23
	Interaction	0.48	2.64	184.47	.676	-
Extraneous cognitive load	Group	0.01	1.00	70.00	.910	-
	Time	9.12	2.58	180.40	< .001	0.12
	Interaction	1.78	2.58	180.40	.161	-
Germane cognitive load	Group	0.37	1.00	70.00	.545	-
	Time	0.39	2.68	187.93	.740	-
	Interaction	0.97	2.68	187.93	.403	-
Total cognitive load	Group	1.45	1.00	70.00	.233	-
	Time	19.16	2.62	183.20	< .001	0.22
	Interaction	0.78	2.62	183.20	.492	-

The preliminary linear regression, see Table 6 (left; iteration 1), revealed a significant contribution only for subject knowledge pretest score and perceived competence. Therefore, further potential predictors were dropped from the model before a second linear regression was carried out, see Table 6 (right; iteration 2). The resulting model represents a fit of the posttest score of students' subject knowledge that accounts for about 50% of the variance in the population.

Prior knowledge (subject knowledge in the pretest) and perceived competence were predictors for subject knowledge in the posttest. Hence, pretest subject knowledge was considered as a covariate in subsequent comparative studies of the AR and non-AR groups. Since perceived competence was assessed as a treatment-dependent variable during the experiment, it could not be considered as an additional covariate. The fact that perceived competence is correlated to posttest performance is rather considered a
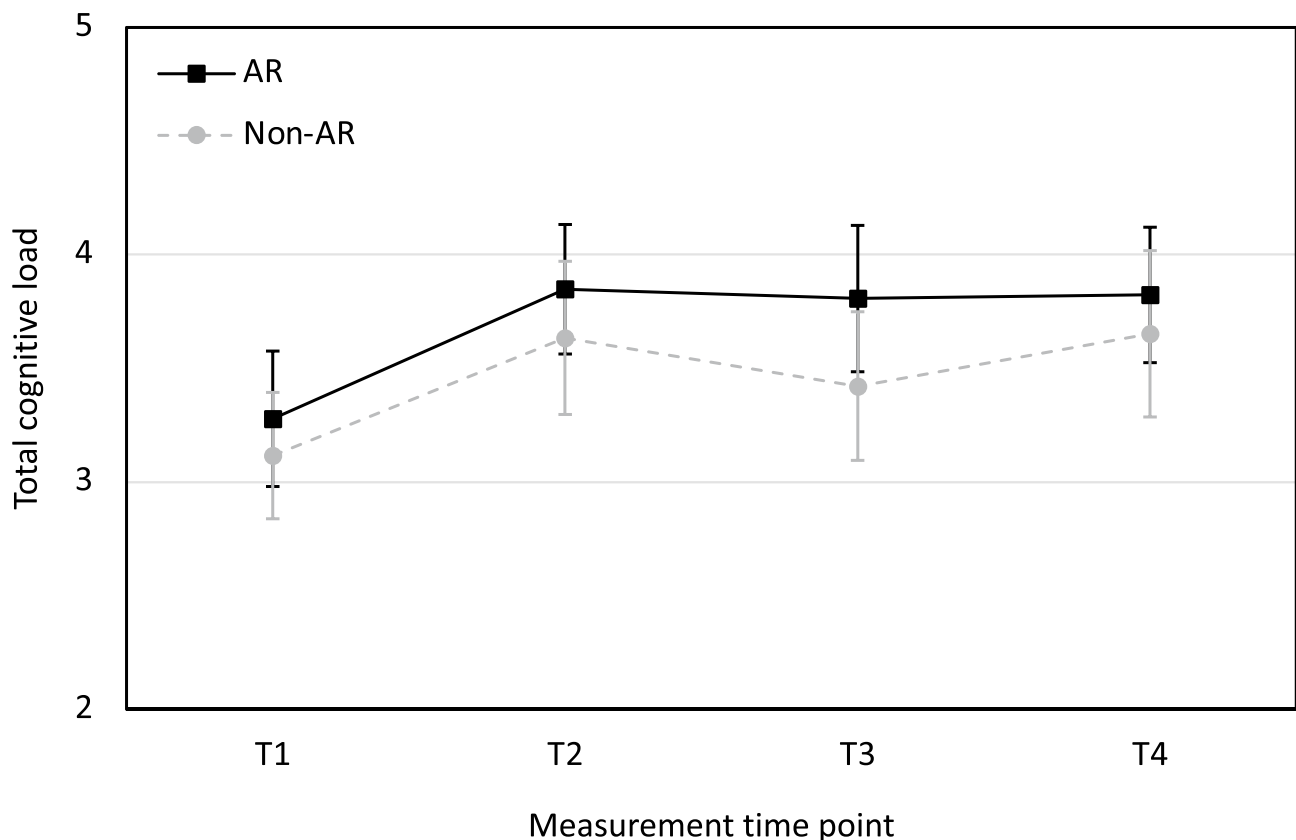
Fig. 7 Time course of total cognitive load over the four measurement time points for the AR and the non-AR groups

Table 6 Fit parameters obtained from regression models for subject knowledge (post) as dependent variable. Parameters with $p > 0.05$ in the first regression were omitted in the second regression

Variable	Iteration 1			Iteration 2		
	$\beta(SD)$	t	p	$\beta(SD)$	t	p
(Intercept)	-	-0.81	.423	-	-0.96	.340
Subject knowledge (pre)	0.55(0.38)	6.17	<.001	0.33(0.37)	6.47	<.001
Interest/enjoyment	0.00(0.70)	0.01	.995	-	-	-
Perceived competence	0.37(0.85)	4.00	<.001	0.30(0.78)	4.27	<.001
Enthusiasm for technology	0.08(0.83)	0.76	.448	-	-	-
Competence in using tech	-0.07(0.89)	-0.59	.558	-	-	-
Adjusted R^2	0.522			0.504		

proof of methodological consistency, because it indicates that the students' self-assessment of their performance during the experiment matched the objective outcome in terms of posttest scores.

A comparison of subject knowledge in the posttest (as dependent) for the AR and non-AR groups was performed using a type II ANCOVA considering subject knowledge in the pretest, perceived competence, and extraneous cognitive load as covariates. The Bonferroni-Holm correction was applied for multiple testing.

The analysis revealed that students in the AR group had a higher subject knowledge on the posttest ($M = 9.57$, $SD = 3.94$) than students in the non-AR group ($M = 8.63$, $SD = 3.46$) (Fig. 8). However, this difference is not significant, $F(1.00, 68.00) = 0.90$, $p = .347$).

The learning achievement was further analyzed with respect to the increase in subject knowledge between pretest and posttest, by performing a type II mixed ANOVA with subject knowledge as a dependent variable, group (AR, non-AR) as a between-subjects factor and time (pretest or posttest) as a within-subjects factor. Only the 8 items that were asked at both test times were selected, see "Learning Achievement" section, to be able to compare the results of the pretest and posttest.

The results (Fig. 9), show a significant main effect with large effect size, $F(1.00, 70.00) = 24.59$, $p < .001$, partial $\eta^2 = 0.26$; we hence conclude that the students overall improved
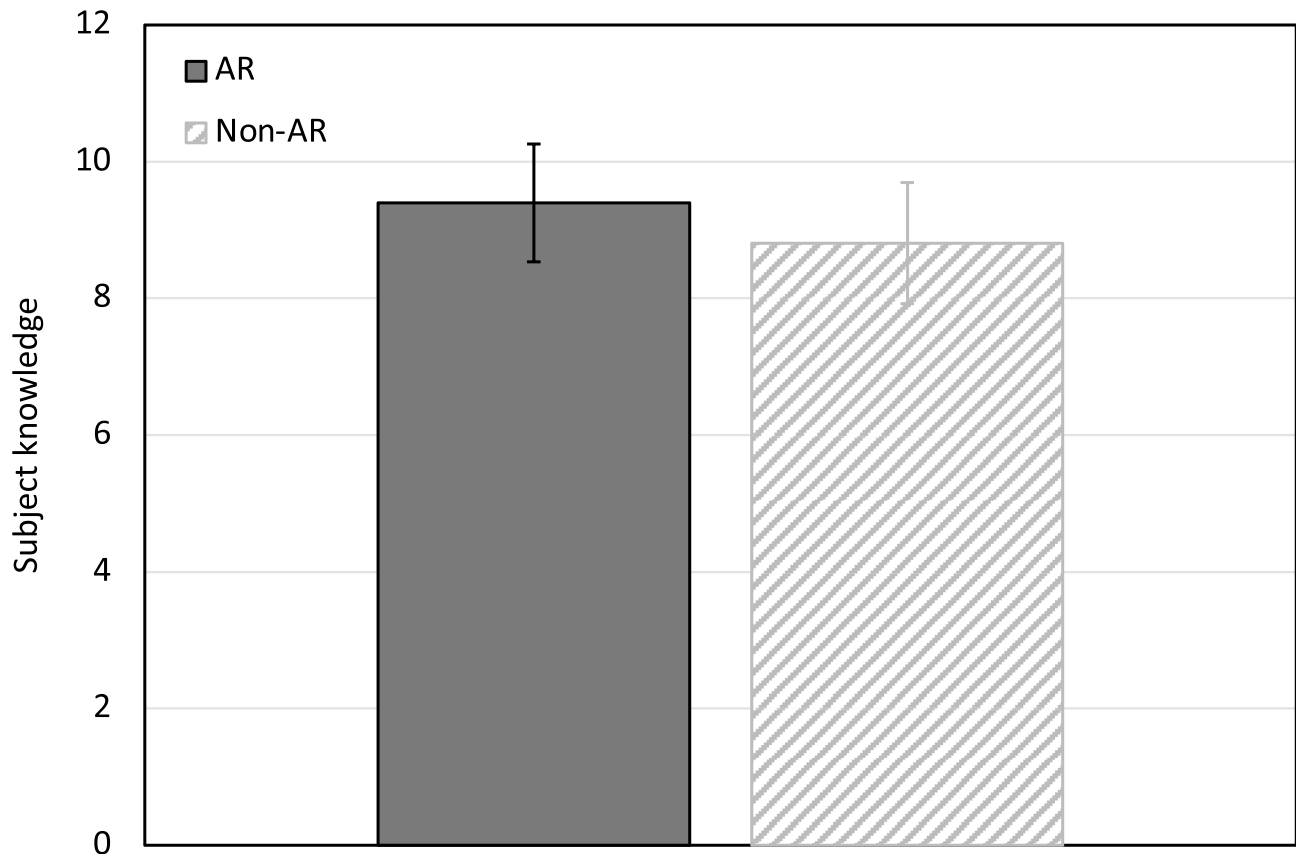
Fig. 8 Estimated marginal means of posttest subject knowledge in the AR group and non-AR group
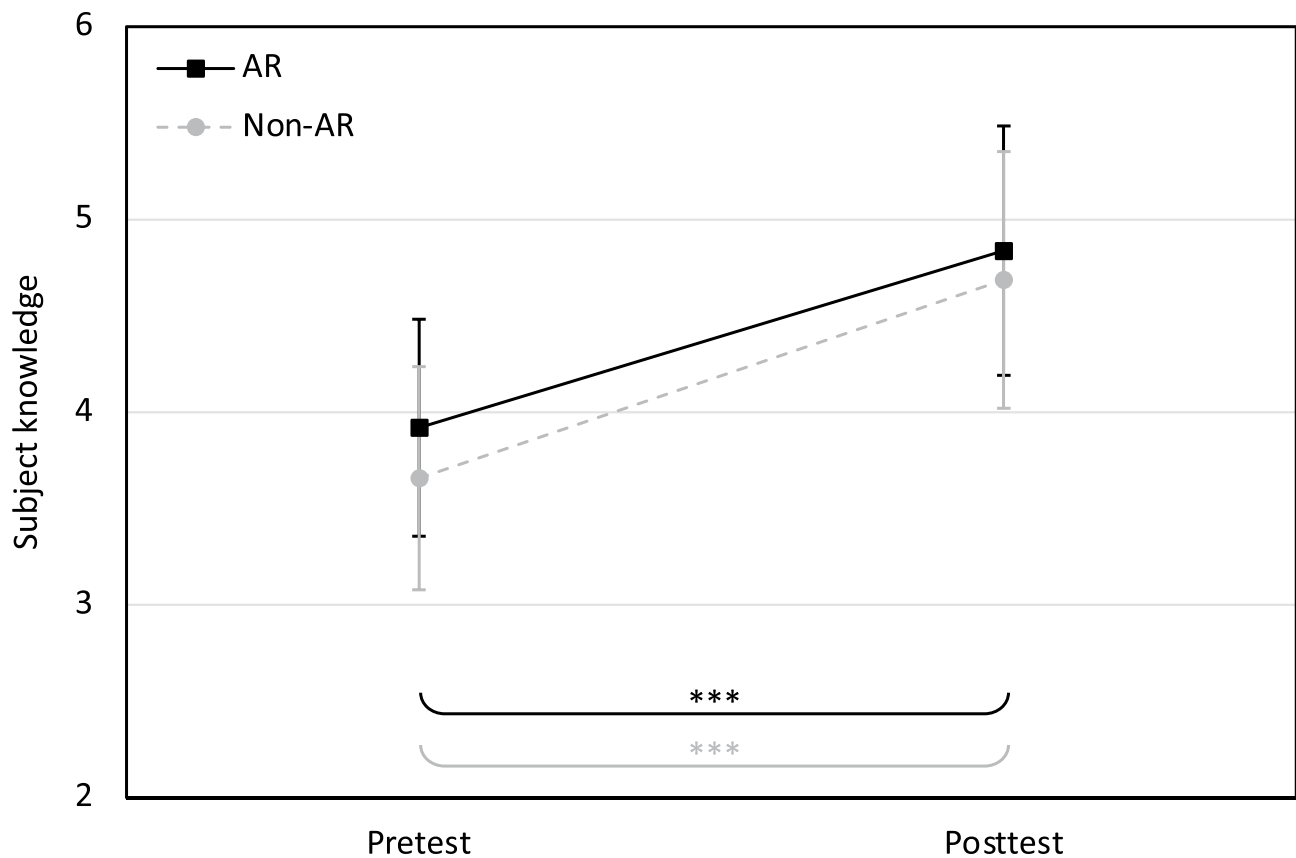
Fig. 9 Interaction plot for AR group and non-AR group and both measurement time points (pretest, posttest)

their performance in the subject knowledge test over the course of the laboratory experiment. However, the interaction effect does not appear to be significant, $F(1.00, 70.00) = 0.08$, $p = 0.781$, meaning that the learning gains in the AR group and the non-AR group are of similar magnitude.

Discussion

For motivation (RQ1), our study shows that learning with AR has advantages over the traditional learning setting for laboratory experiments. The small to moderate overall effect ($p = .039$, partial $\eta^2 = 0.06$) can be traced back to large effects in interest/enjoyment ($p = .002$, partial $\eta^2 = 0.14$) in working with the AR environment, but these large effects were countered by the absence of differences in perceived competence. This result appears consistent with the slightly heterogeneous preliminary findings on the use of AR in learning settings in general (Cao & Yu, 2023; Chang et al., 2022) and for science education in particular (Ibáñez & Delgado-Kloos, 2018 However, the present data go beyond previous research as they report findings for the use of AR glasses as a technology, which have not been analyzed before. The fact that perceived competence of learners did not differ significantly

across groups is surprising because only about half of the learners in the AR group reported prior experience with immersive headsets (19 of 39 learners).

For cognitive load (RQ2), the only significant difference between the AR group and non-AR group was found for intrinsic cognitive load ($p = .042$, partial $\eta^2 = 0.06$). This difference might have occurred due to differences, for example, in learners' prior knowledge. Another possibility is that the additional visualizations were perceived as increasing the complexity of the learning environment because learners in the AR group were required to make sense of these representations to successfully complete the tasks. While learners in the non-AR condition also had the visualizations available in their workbooks, they did not necessarily need to reflect on them to read out measurement values from the multimeter. Also surprising was that no detectable differences were found in extraneous cognitive load between the groups. Based on the Cognitive Theory of Multimedia Learning (Mayer, 2009) and a broad empirical basis on the spatial contiguity effect (Ginns, 2006; Schroeder & Cenkci, 2018), one might have assumed that the spatial integration of the experimental components and the setup as well as the model-related visualizations in the AR group should have reduced the extraneous cognitive load. Yet, most studies on

AR use have not shown that AR clearly reduces the cognitive load in learning situations in general (Radu, 2014; Buchner et al., 2022). Specifically in science education, studies show no clear trend regarding cognitive load for AR (Altmeyer et al., 2020; Elford et al., 2022; Thees et al., 2020). However, in this context, one might argue that findings in which both groups have equal amounts of cognitive load represent a benefit of the AR environment, because dealing with a new technology can be considered an additional challenge that is absent for learners in the non-AR environment (Radu, 2014).

The most striking finding in relation to previous research was for the students' learning achievement (RQ3). Although students in the AR group showed a higher subject knowledge on the posttest ($M = 9.57$, $SD = 3.94$) than students in the non-AR group ($M = 8.63$, $SD = 3.46$), there was no significant difference in learning growth in subject knowledge ($p = .347$). Both for the application of AR in education in general (Malone et al., 2023) and for the special field of science education and specifically for physics as a discipline (Xu et al., 2022), clear findings in systematic reviews and meta-analyses have shown learning benefits associated with AR compared to traditional learning settings. The present data appear to support the findings of Thees et al. (2020) and, thereby, open the question of how the specific application of AR glasses differs from other fields of application of AR and other technologies. As reported earlier, our data showed that the perceived competence of both groups was not significantly different, suggesting that the lack of advantage in the AR group was not due to difficulties in using the technology.

In addition to the specific results comparing AR to non-AR, it seems interesting that the laboratory course in this study proves also suitable for training further content knowledge in both groups, as there was a strong overall effect in learning achievement ($p < .001$, partial $\eta^2 = 0.26$). This appears to contradict popular analyses from the field of physics education (Holmes & Wiemann, 2018; Holmes et al., 2020) and raises the question of the extent to which lab courses are suitable for the development of content knowledge through the integration of concept-related content, including visualizations. A possible explanation may be found by Xu et al. (2022) indicating strong advantages of AR in the field of earth sciences can be traced back to this field's specific focus on spatial abilities. In contrast, AR has shown no advantages in biology, where it is assumed that knowledge of real objects is fundamental. For the learning setting of the present study, it can be assumed that spatial abilities are not particularly important for learning the concept of optical polarization; but what is important is the integration of AR with real experimental setups. Instead of focusing on features of the AR environment to explain the absence of effects, an interesting alternative explanation might instead to consider positive features of the design of

the non-AR environment, which was explicitly designed to provide a fair comparison condition. This included providing the students with the same visualizations (although static and not interactive) and prompting them to generate their own visualizations of measurement data. Hence, the non-AR environment was designed to explicitly compensate for potential drawbacks based on the differences in technological implementation. This explanation highlights the primacy of good instructional design independent of technological implementation (Feldon et al., 2021), and it further points out the importance of establishing fair conditions in media comparison studies to obtain meaningful results.

Research Limitations

This study was designed with special consideration for enabling a fair comparison between AR and non-AR conditions, particularly with respect to the available representations. Critically, however, we cannot exclude that novelty effects due to the experience of using a new technology (Hamari et al., 2014) might have positively influenced learners' motivation in the AR group. This is an issue that has also not been rigorously accounted for in previous studies related to educational technology in general (Tsay et al. 2019).

In addition, this study focused on quantitative measures such as learning achievement and cognitive load. To gain a deeper understanding of the quantitative results, it would be desirable to investigate how learners interact and work with the experiments from a qualitative perspective. For example, it is not yet clear how learners in the AR and non-AR environments perceived the provided representations and how they included them in their learning process, or how the interactive functions of the AR environment might have influenced their experimental strategies.

The unequal distribution of female and male learners results in a limitation for the present study regarding the sample, as the proportion of female participants overall was only just over 20 percent. This proportion is roughly comparable to the likewise significantly lower proportion of females compared to males in the study by Thees et al. (2020) and corresponds to the typical gender gap of women within STEM education (Cimpian et al., 2020). The extent to which the unequal distribution influences the results of the study appears unclear, as both Ibáñez and Delgado-Kloos (2018) and Xu et al. (2022) point out that the influence of gender on learning in AR-based learning environments in science education has not yet been investigated.

A further limitation relates to the selection of the survey instruments. The survey for cognitive load was carried as a self-assessment, which is quite common, but more objective measurement methods would be desirable. As this was a laboratory experiment, it also seems relevant to examine the extent

to which the experimental approach itself influenced the results; this could be done, for example, by checking whether the experimental tasks were performed similarly well in both comparison groups. For both aspects mentioned above, it seems sensible to incorporate eye-tracking data, which can easily be done using AR glasses (Souchet et al., 2022).

Conclusion and Implications

We examined the extent to which the use of AR glasses in laboratory experiments is beneficial for students' motivation, cognitive load and learning achievement compared to a traditional learning setting. We used an experimental comparison group design with $N = 75$ students in the context of a laboratory experiment on optical polarization.

In line with existing findings, the results show that learners appear more motivated when using an AR application. However, a novelty effect cannot be excluded, so longitudinal surveys over a longer intervention period seem necessary. Regarding cognitive load, no significant difference between learners with and without AR were found. On the one hand, this is promising, as potential challenges in dealing with AR as a novel technology have not proved to be a hinderance (Buchner & Kerres, 2023). But, on the other hand it is also surprising, as the significantly higher (spatial) contiguity in the AR experiment could be expected to lower the extraneous cognitive load, according to theoretical and empirical findings on the Cognitive Theory of Multimedia Learning. Finally, it should be noted that contrary to most studies (Bacca et al., 2014; Cao & Yu, 2023; Chang et al., 2022; Zhang et al., 2022), all learners acquired knowledge in the same way regardless of the learning technology. This contradicts the findings from meta-analyses on the use of AR in physics (Xu et al., 2022) but is consistent with the only finding to date (according to the present literature review of our study) on the use of AR glasses for learning in physics (Thees et al., 2020). As such, these results could suggest that it may be inappropriate for media comparison studies to focus on surface features, such as the specific AR technology or the teaching subject. Rather, future studies should focus on the content design of AR applications and, by analyzing the deep structure, identify features that make AR applications conducive to learning. A good starting point for future research could be to examine the initial findings of existing comparative studies on AR in biology applications, which showed real experiences were more important than virtual ones, and studies on AR in earth sciences applications, which showed that AR was particularly supportive in developing spatial ability skills critical for the field. These findings should be investigated more systematically and transferred to other subjects such as physics.

Data Availability The data and materials of the study are not publicly accessible due to the written consent given by the subjects to participate.

Declarations

Ethical Approval Not applicable.

Informed Consent The consent of the individual participants in form of their consent to participate and their consent to publish is described in the respective sections.

Consent to Participate Written informed consent was obtained from all individual participants included in the study.

Consent to Publish The authors affirm that human research participants provided informed consent for publication of the study.

Statement Regarding Research Involving Human Participants and/or Animals The study examines human participants. Before participating in the study, they were informed about the conditions of the study as well as the handling of data and the preservation of anonymity.

Competing Interests Not applicable.

References

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education, 166*, 104154. https://doi.org/10.1016/j.compedu.2021.104154

Altmeyer, K., Kapp, S., Thees, M., Malone, S., Kuhn, J., & Brünken, R. (2020). The use of augmented reality to foster conceptual knowledge acquisition in STEM laboratory courses – Theoretical background and empirical results. *British Journal of Educational Technology, 51*, 611–628. https://doi.org/10.1111/bjet.12900

Antonietti, A., Rasi, C., Imperio, E., & Sacco, M. (2000). The representation of virtual reality in education. *Education and Information Technologies, 5*, 317–327. https://doi.org/10.1023/A:1012057608694

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*, 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*, 34–47. https://doi.org/10.1109/38.963459

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. *Educational Technology & Society, 17*, 133–149.

Backhaus, J., Huth, K., Entwistle, A., Homayounfar, K., & Koenig, S. (2019). Digital affinity in medical students influences learning outcome: A cluster analytical design comparing vodcast with traditional lecture. *Journal of Surgical Education, 76*, 711–719. https://doi.org/10.1016/j.jsurg.2018.12.001

Binder, J. S., Scholz, M., Ellmann, S., Uder, M., Grützmann, R., Weber, G. F., et al. (2021). Cinematic rendering in anatomy: A crossover study comparing a novel 3D reconstruction technique to conventional computed tomography. *Anatomical Sciences Education, 14*, 22–31. https://doi.org/10.1002/ase.1989

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning, 38*, 285–303. https://doi.org/10.1111/jcal.12617

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education, 195*, 104711. https://doi.org/10.1016/j.compedu.2022.104711

Cai, Y., Pan, Z., & Liu, M. (2022). Augmented reality technology in language learning: A meta-analysis. *Journal of Computer Assisted Learning, 38*, 929–945. https://doi.org/10.1111/jcal.12661

Cao, W., & Yu, Z. (2023). The impact of augmented reality on student attitudes, motivation, and learning achievements – a meta-analysis (2016–2023). *Humanities & Social Sciences Communications, 10*, 352. https://doi.org/10.1057/s41599-023-01852-2

Chang, H.-Y., Binali, T., Liang, J.-C., Chiou, G.-L., Cheng, K.-H., Lee, S.W.-Y., et al. (2022). Ten years of augmented reality in education: A meta-analysis of (quasi-)experimental studies to investigate the impact. *Computers & Education, 191*, 104641. https://doi.org/10.1016/j.compedu.2022.104641

Cimpian, J. R., Kim, T. H., & McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science, 368*, 1317–1319. https://doi.org/10.1126/science.aba7377

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Conole, G., & Dyke, M. (2004). What are the affordances of information and communication technologies? *ALT-J Research in Learning Technology, 12*, 113–124. https://doi.org/10.1080/0968776042000216183

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Csikszentmihalyi, M. (1975). Play and unique prizes. *Journal of Humanistic Psychology, 15*, 41–63. https://doi.org/10.1177/002216787501500306

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper-Perennial.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*, 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*, 7–22. https://doi.org/10.1007/s10956-008-9119-1

Elford, D., Lancaster, S. J., & Jones, G. A. (2022). Exploring the effect of augmented reality on cognitive load, attitude, spatial ability, and stereochemical perception. *Journal of Science Education and Technology, 31*, 322–339. https://doi.org/10.1007/s10956-022-09957-0

Feldon, D. F., Jeong, S., & Clark, R. E. (2021). Fifteen common but questionable principles of multimedia learning. In R. E. Mayer, & L. Fiorella (eds.), *The Cambridge handbook of multimedia learning* (pp. 25–40). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108894333.005

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

Hamari, J., Koivisto, & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In R. H. Sprague, Jr. (eds.), *47th Hawaii International Conference on System Sciences* (pp. 3025–3034). Los Alamitos: IEEE Computer Society. https://doi.org/10.1109/HICSS.2014.377

Holmes, N. G., Keep, B., & Wiemann, C. E. (2020). Developing scientific decision making by structuring and supporting student agency. *Physical Review Physics Education Research, 16*, 010109. https://doi.org/10.1103/PhysRevPhysEducRes.16.010109

Holmes, N. G., & Wiemann, C. E. (2018). Introductory physics labs: We can do better. *Physics Today, 71*, 38–45. https://doi.org/10.1063/PT.3.3816

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education, 123*, 109–123. https://doi.org/10.1016/j.compedu.2018.05.002

Ibáñez, M. B., Di Serio, Á., Villarán, D., & Delgado Kloos, C. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education, 71*, 1–13. https://doi.org/10.1016/j.compedu.2013.09.004

Jin, F., & Divitini, M. (2020). Affinity for technology and teenagers' learning intentions. In A. Robins, A. Moskal, A. J. Ko, & R. McCauley (eds.), *Proceedings of the 2020 ACM conference on international computing education research* (pp. 48–55). New York: Association for Computing Machinery. https://doi.org/10.1145/3372782.3406269

Kapp, S., Thees, M., Strzys, M. P., Beil, F., Kuhn, J., Amiraslanov, O., et al. (2019). Augmenting Kirchhoff's laws: Using augmented reality and smartglasses to enhance conceptual electrical experiments for high school students. *The Physics Teacher, 57*, 52–53. https://doi.org/10.1119/1.5084931

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen – der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), *Der Mensch im Mittelpunkt technischer Systeme* (pp. 196–201). VDI.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Laun, M., Czech, C., Hartmann, U., Terschüren, C., Harth, V., Karamanidis, K., et al. (2022). The acceptance of smart glasses used as side-by-side instructions for complex assembly tasks is highly dependent on the device model. *International Journal of Industrial Ergonomics, 90*, 103316. https://doi.org/10.1016/j.ergon.2022.103316

Malone, S. Garzón, J., & Kuhn, J. (2023).Three decades of augmented reality in education: A second-order meta-analysis and research synthesis. OSF Preprints. https://doi.org/10.31219/osf.io/amw4t

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity. In R. E.

Mayer (eds.), *The Cambridge handbook of multimedia learning* (pp. 279–315). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015

Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 69*, 48–58. https://doi.org/10.1080/02701367.1989.10607413

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems, 77*, 1321–1329.

Mills, L., Knezek, G., Tyler-Wood, R., Christensen, R., & Dunn-Rankin, P. (2013). Exploring the relationship between middle school student technology affinity and attitudes toward school. In R. McBridge, & M. Searson (eds.), *Proceedings of SITE 2013 – Society for Information Technology & Teacher Education International Conference* (pp. 2269–2272). New Orleans: Association for the Advancement of Computing in Education. Retrieved from: https://www.learntechlib.org/p/48445/

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal Ubiquitous Computing, 18*, 1533–1543. https://doi.org/10.1007/s00779-013-0747-y

Schlummer, P., Abazi, A., Borkamp, R., Lauströer, J., Schulz-Schaeffer, R., Schuck, C., Pernice, W., Heusler, S., & Laumann, D. (2023). Seeing the unseen - enhancing and evaluating undergraduate polarization experiments with interactive Mixed-Reality technology. *European Journal of Physics, 44*, 065701. https://doi.org/10.1088/1361-6404/acf0a7

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review, 30*, 679–701. https://doi.org/10.1007/s10648-018-9435-9

Siebert, F., Roesler, E., & Karrer-Gauß, K. (2022). Validating the TA EG questionnaire in germany and comparing affinity for technology between Japan, Mexico, and the USA. *PsychArchives*. https://doi.org/10.23668/psycharchives.7056

Sırakaya, M., & Sırakaya, D. A. (2020). Augmented reality in STEM education: A systematic review. *Interactive Learning Environments, 8*, 1556–1569. https://doi.org/10.1080/10494820.2020.1722713

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human-Computer Interaction, 38*, 801–824. https://doi.org/10.1080/10447318.2021.1976509

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology, 60*, 100832. https://doi.org/10.1016/j.newideapsych.2020.100832

Strzys, M. O., Kapp, S., Thees, M., Klein, P., Lukowicz, P., Knierim, P., et al. (2018). Physics holo.lab learning experience: Using smartglasses for augmented reality labwork to foster the concepts of heat conduction. *European Journal of Physics, 39*, 035703. https://doi.org/10.1088/1361-6404/aaa8fb

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (eds.), *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Amsterdam: Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Szalavári, Z., Schmalstieg, D., Fuhrmann, A., & Gervautz, M. (1998). "Studierstube": An environment for collaboration in augmented reality. *Virtual Reality, 3*, 37–48. https://doi.org/10.1007/BF01409796

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior, 108*, 106316. https://doi.org/10.1016/j.chb.2020.106316

Tsay, C.H.-H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2019). Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning, 36*, 128–146. https://doi.org/10.1111/jcal.12385

Wagner, T., Hoyer, C., Ringl, C., & Kuhn, J. (2023). Investigating diffraction phenomena with low-cost material and augmented reality. *The Physics Teacher, 61*, 402–403. https://doi.org/10.1119/5.0149766

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Testing a short scale of intrinsic motivation. *Zeitschrift für Didaktik der Naturwissenschaften, 15*, 31–45.

Xu, W.-W., Su, C.-Y., Hu, Y., & Chen, C.-H. (2022). Exploring the effectiveness and moderators of augmented reality on science learning: A meta-analysis. *Journal of Science Education and Technology, 31*, 621–637. https://doi.org/10.1007/s10956-022-09982-z

Zhang, J., Li, G., Huang, Q., Feng, Q., & Luo, H. (2022). Augmented reality in K-12 education: A systematic review and meta-analysis of the literature from 2000 to 2020. *Sustainability, 14*, 9725. https://doi.org/10.3390/su14159725