Coban et al. *EPJ Quantum Technology* (2025) 12:15
https://doi.org/10.1140/epjqt/s40507-025-00310-z

EPJ Quantum Technology
a SpringerOpen Journal

RESEARCH Open Access

AI support meets AR visualization for Alice and Bob: personalized learning based on individual ChatGPT feedback in an AR quantum cryptography experiment for physics lab courses

Atakan Coban[1*], David Dzsotjan[1], Stefan Küchemann[1], Jürgen Durst[2], Jochen Kuhn[1] and Christoph Hoyer[1]

*Correspondence:
atakan.coban@lmu.de
[1]Chair of Physics Education, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU), Munich, Germany
Full list of author information is available at the end of the article

Abstract

Quantum cryptography is a central topic in the quantum technology field that is particularly important for secure communication. The training of qualified experts in this field is necessary for continuous development. However, the abstract and complex nature of quantum physics makes the topic difficult to understand. Augmented reality (AR) allows otherwise invisible abstract concepts to be visualized and enables interactive learning, offering significant potential for improving quantum physics education in university lab courses. In addition, personalized feedback on challenging concepts can facilitate learning, and large language models (LLMs) like ChatGPT can effectively deliver such feedback. This study combines these two aspects and explores the impact of an AR-based quantum cryptography experiment with integrated ChatGPT-based feedback on university students' learning outcomes and cognitive processes. The study involved 21 groups (11 Group A; 10 Group B) of students in a physics laboratory course at a German university and used four open-ended questions to measure learning outcomes and gaze data as a learning process assessment. Statistical analysis was used to compare scores between feedback and non-feedback questions, and the effect of ChatGPT feedback on eye-tracking data was examined. The results show that ChatGPT feedback significantly improved learning outcomes and affected gaze data. While the feedback on conceptual questions tended to direct attention to the visualizations of the underlying model, the feedback on questions about experimental procedures increased visual attention to the real experimental materials. Overall, the results show that AI-based feedback draws visual attention towards task-relevant factors and increases learning performance in general.

Keywords: Quantum physics education; Augmented reality; Large language models; Cognitive feedback; Cognitive learning outcome; Cognitive learning process; Eye-tracking

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

1 Introduction

Quantum technology, which is based on the fundamental principles of quantum mechanics, is driving revolutionary advancements in sectors from computing to telecommunications. Quantum computing technologies utilize the principles of quantum physics to address complex problems ranging from cryptography to optimization and quantum simulation by performing calculations at speeds that are unattainable by classical computers [1]. In particular, integrated quantum photonics improves the generation and control of quantum states at the chip scale, setting new boundaries for quantum communication and information processing applications [2]. This technology not only improves the manipulation and stability of quantum states but facilitates more efficient and secure communication networks [3].

Quantum cryptography, which is integral to secure communication, is a very important and rapidly developing quantum technology [4]. It is based on fundamental concepts from quantum physics, such as its probabilistic nature, the polarization state of a single photon, and the principle of superposition. Therefore, integrating quantum cryptography effectively into real-life applications requires a high level of quantum physics knowledge and application skills. Intensive, effective quantum physics education is the primary way to achieve this.

The integration of quantum technologies, such as quantum cryptography, into existing telecommunications infrastructure presents various obstacles. One of the most significant current challenges is the shortage of human resources who can effectively manage quantum communication infrastructures, which is defined in the literature as the 'quantum skills gap' [5]. One of the main reasons for this gap is that quantum physics education has fallen behind technological advancements. Therefore, the primary and most crucial method to address this gap is to include comprehensive quantum physics training in education [6, 7].

One approach to imparting knowledge about quantum cryptography is a physics lab course. As quantum cryptography is an extremely abstract and complicated topic, learners could be provided with virtual visualizations of quantum states via AR within the in-person setup. In this paper, we adopt this approach and examine how much individualized feedback provided by ChatGPT can improve the answers of participants interacting with an AR quantum cryptography experiment. Eye tracking data analysis provided additional insights into the learners' 3D visual attention during the lab course and the effects of verbal ChatGPT feedback on their gaze.

The integration of ChatGPT-based feedback into an AR-based quantum physics experiment and the analysis of the effects of that feedback in terms of learning outcome and learning process using gaze data are the key elements that make this study significant.

2 Theoretical background

2.1 Quantum cryptography: a brief overview

Methods for encrypting and sending sensitive information between two parties so that a third party cannot access the content have been pursued in various human cultures and societies since the dawn of human history. Practically, all of these methods or protocols have two key points in common: 1) a key or rule is used to encrypt the message on the transmitting side and decrypt it on the receiving side, and 2) both sides must possess the key, but no one else can have it. The security of the protocol depends on the strength of

the encryption and the likelihood that no one other than the sending and receiving parties can acquire the encryption key.

Next, we will describe the BB84 protocol, a key distribution method that relies on the laws of quantum mechanics and is, in principle, perfectly secure.

2.1.1 One-time pad encryption

At the heart of the BB84 protocol lies a classical method for transmitting an encrypted message: the one-time pad encryption, or single-use key encryption, method. The principle can be demonstrated with a simple example. Suppose we want to encrypt a message that is the letter A. We add its binary ASCII encoding and a randomly generated binary key (mod1 addition) to obtain the encrypted message:

$$\begin{aligned}&\text{message (A): 01000001, key: 01010110}\\&\text{encrypted message: 01000001 + 01010110 = 00010111,}\end{aligned} \tag{1}$$

which can then be decoded on the receiving side by adding the encryption key to it again:

$$\text{recovered message: 00010111 + 01010110 = 01000001} \tag{2}$$

resulting in the message that we wanted to encrypt. The encryption key must be used one time only, i.e., for each message string, a new random key must be generated. This classical method, in principle, is 100% secure, provided that only the sending and the receiving party have the encryption key. The problem of secure data transmission is therefore reduced to transmitting an encryption key from the sender to the receiver such that no third party can intercept it. The use of quantum bits within the BB84 protocol and the laws of quantum mechanics make this possible and even permit the presence of a potential eavesdropper to be detected.

2.1.2 Key distribution and BB84 protocol

In the BB84 protocol, we surpass mere 1s and 0s and introduce the concept of bases. This is conveniently demonstrated in our experiment, where polarized photons serve as quantum bits (qubits). With lambda/2 plates, the polarization of incoming linearly polarized light can be rotated at will to generate an outgoing photon of arbitrary linear polarization. Consider two sets of photonic polarization basis states:

$$\{|H\rangle, |V\rangle\}, \{|+45\rangle, |-45\rangle\}, \tag{3}$$

where the first pair are the horizontal and vertical polarization states and the second are the states of polarization in the +45- and −45-degree directions, respectively. We name the bases + and x, respectively, and define two sets of qubits as

$$\begin{aligned}&|0\rangle_+ = |H\rangle, |1\rangle_+ = |V\rangle,\\&|0\rangle_X = |-45\rangle = 1/\sqrt{2}(|H\rangle - |V\rangle),\\&|1\rangle_X = |+45\rangle = 1/\sqrt{2}(|H\rangle + |V\rangle)\end{aligned} \tag{4}$$

Thus, the transmitted information has two layers: the value of the bit (0 or 1) and the value of the base (+ or x).

The quantum key distribution within the BB84 protocol happens as follows. The sending party (Alice), by rotating their lambda/2 plate, generates photons with random bases (+ or x) and bit values (0 or 1). The receiving party (Bob), in turn, rotates their own lambda/2 plate, randomly choosing whether to measure in the + or the x basis, and measures an incoming string of 0s and 1s. For instances where both Alice and Bob set their bases to the same value, then a 1 sent by Alice will always be measured as a 1 by Bob. However, if they set different bases, say, Alice sends $|1\rangle_X$ and Bob's measuring device is set in the + base, then, according to Eq. (4), this incoming photon on Bob's side will be randomly mapped either to the 0 or the 1 of the + base. Crucially, as single photons are being sent, Bob's detectors never click simultaneously: He either measures a 0 or a 1. The correspondents can be sure that Bob's measurements were correct when his choice of base coincides with Alice's choice of base, e.g., a 1 sent by Alice was measured as a 1 by Bob. Eventually, both Alice and Bob will have a list of base values and bit values. As the last step of the quantum key distribution, they use a classical information channel to compare their base sequences but not their bit values. The qubit transmissions where their bases did not match are discarded and the remaining string of 0s and 1s serve as the shared key ready to be used in the encryption.

2.1.3 Detection of an eavesdropper

The detection of the presence of an eavesdropper is where quantum key distribution truly shines. Suppose that a third person, Eve, can capture the stream of photons sent by Alice. As the no-cloning theorem states, Eve cannot read out the photon sent by Alice and simultaneously send an identical photon to Bob: Eve must generate new photons pass on to Bob. Therefore, Eve has both a detector and an emitter that work in both + and x bases and can measure and generate 0s and 1s in these respective bases.

With Eve present, even when Alice's and Bob's bases coincide (both either + or x), their bit values will not always be equal! As Table 1 shows, at every instance when Eve's emitter base varies from Alice's and Bob's bases, there is a 50% chance that Bob measures a different bit value than what Alice sent. As the base choices are completely random and Alice's and Bob's bases are equal, in an average of 50% of the cases, Eve's base will not match Alice's and Bob's (see Table 1). In cases of a base mismatch between Eve and the other two, there is a further 50% chance that Bob will measure a different bit value than that sent by Alice. Therefore, assuming Alice's and Bob's bases match, the overall probability of wrong bit transmission is the product of these chances, i.e., 25%.

To detect the potential eavesdropper, after comparing the bases and discarding the irrelevant instances, Alice and Bob also publicly compare some of the remaining bit values.

Table 1 Probability of matching bit values for Alice and Bob, for several different base choices, provided Alice and Bob choose the same base. Eve's base will not match Alice's and Bob's in about 50% of the cases

Alice's and Bob's bases	Eve's emission base	Probability that Alice's and Bob's bit values match
++	+	100%
++	X	50%
XX	+	50%
XX	X	100%

If they see a bit value mismatch in about 25% of the cases, they can reliably state that security has been compromised and discard the quantum key. This happens before any part of the message itself has been transmitted, giving Eve no way of accessing it.

2.2 Physics education research on quantum cryptography

Quantum physics concepts cannot be directly observed due to their abstract nature [8, 9]. The mathematical formalism of quantum mechanics adds another layer of complexity to the topic [10]. These factors make it difficult for students to contextualize quantum physics concepts and relate them to real-world phenomena, creating significant barriers to the learning process [11].

In particular, learning quantum cryptography requires cognitive skills and practical application of the related concepts, in addition to the challenges of learning quantum physics, further complicating the learning process [12]. Effective learning in this field entails visualizing abstract concepts to ensure students' active participation in the process and show where these concepts can be used in real life through practical processes [13]. Therefore, supporting quantum cryptography education with visual tools and interactive methods is believed to improve learning outcomes; as a result, hands-on experiments, games and interactive simulations have become very popular for teaching quantum cryptography [14–17].

Vadla et al. [18] reported that QuaSim [15], a gamified intelligent tutoring system developed for teaching quantum cryptography, had positive effects on students' achievement and interest in processes involving their active participation. Utama et al. [19] allowed students at the pre-university level to implement the BB84 protocol with the experiment set they developed for the high-school level and received feedback from students. The students expressed satisfaction with the application and appreciated the opportunity to practice the concepts they had learned. Pallotta [20] stated that the QuVis [17] simulation was greatly effective in supporting students in analysing the different elements of the key distribution process and that students evaluated this type of activity very highly as it helped them understand how quantum physics is not only an abstract or intuitive theory but also a technological resource.

An activity focused on understanding how to build a secure protocol with quantum cryptography applications has dual advantages of raising awareness about the potential of quantum technologies as well as creating a learning context to deepen students' understanding of quantum superposition and quantum measurement [20]. Such activities allow students to consolidate their theoretical knowledge in practice and may both raise their awareness of quantum technologies and help them to better grasp the basic principles of quantum mechanics [21].

Therefore, experimental processes involving quantum cryptography are important both for students to have the opportunity to realize quantum applications in the field and learn about quantum concepts by observing how they are used in practice. Thorlabs has developed the *Quantum Cryptography Analogy Demonstration Kit*, an experiment set for the BB84 protocol on quantum cryptography suitable for the high school and university levels [22].

The experimental set allows students to learn how to determine a secret keycode by following the BB84 protocol. Through this collaborative research-based approach, they can gain insights into the probabilistic structure of the quantum world, the disruption of quantum systems by observation, and the idea of data transmission with quantum bits [23].

As the underlying physical concepts are highly complex and difficult for learners to understand, an adapted version of the original setup that provides individual feedback and intuitively understandable real-time representations of the invisible quantum states might be beneficial. In such an adapted version of the experiment, augmented reality glasses would show virtual visualizations of quantum states that would otherwise not be directly observable, making them visible within the real-world setup. Furthermore, an option for verbal interaction with ChatGPT could offer automated and, primarily, individual feedback on students' answers. Such feedback may drive the student group to discuss questions further, improving the accuracy of their answers [24].

The learning method employed in this study combines hands-on experimental activities with virtual visualizations, and ChatGPT for providing feedback to support students' learning outcomes. Before discussing the study in greater detail, an overview of the theories related to learning with multiple representations, augmented reality, and formative feedback will be presented.

2.3 Multiple representations and augmented reality

For scientific learning, representing content through different types of representations, or multiple external representations (MERs), is well-documented as beneficial in the natural sciences [25] and physics [26]. This holds especially true for abstract content, such as quantum physics. MERs are especially important for conceptual understanding [27] and are discussed as a necessary condition for in-depth understanding [28].

For visualizing MERs, technology-enhanced learning environments have often been used, e.g. simulations (for quantum physics; [29]). AR technology combined with real-world experimental setups has become increasingly prevalent in various fields, e.g. to provide 3D visualizations and enhance the real-world environment with overlaid virtual content [30, 31]. AR systems' key elements include the alignment of virtual 3D representations with physical objects in the real world and an interactive 3D visualization that changes in real time based on interactions with these physical objects [32].

There is broad evidence and empirical consensus that AR environments foster learning in general [33, 34]. A recent second-order meta-analysis compared AR-based learning environments to traditional learning environments without AR [35]. The mean effect size can be interpreted such that all individual analyses demonstrate medium to large effects and are, therefore, consistent regarding the positive impact of AR on learning achievement. AR-driven learning activities simplify students' mental processes [30] and enhance their understanding of physical phenomena [36]. In quantum physics laboratory settings, AR technology can provide a more understandable conceptual process by presenting abstract quantum physics concepts that cannot be observed directly as virtual models and allowing users to interact with them [37]. Moreover, AR departs from two-dimensional displays by allowing concepts to be observed and interacted within three dimensions. This shift to spatial and interactive learning can have profound effects on conceptual understanding [38]. Such environments can help students engage with complex quantum phenomena more tangibly and intuitively, enhancing their understanding and involvement in the subject [39]. Therefore, learning about unobservable quantum physics concepts by means of three-dimensional virtual representations connected to the real world via AR and applications in which students can interact with these virtual representations and observe their changes instantaneously can have a positive effect on student success. See Dzsotjan et al. [40] for a detailed discussion of the MER used in this AR environment.

2.4 Formative feedback via large language models

In the field of education, LLMs can be utilized in various beneficial ways, such as delivering lectures, solving examples, providing personalized assistance and developing educational materials [41]. LLMs can provide formative feedback to students [42]. In education, personalized feedback adapted to a student's learning situation is crucial, as it fundamentally supports students in monitoring their progress and making meaningful connections with the material [43]. Feedback in the educational environment improves learning outcomes [44–46] and is of significant importance in quantum physics education. Due to the abstract nature and advanced mathematical structure of quantum physics, identifying and addressing students' confusions and learning difficulties during the learning process is essential, as is asking students questions throughout the process and providing feedback based on their responses to diagnose their current understanding and enhance their learning processes [9]. This is vital for preventing and correcting learning difficulties that may arise from incorrect understandings. The approach creates a more structured and effective educational experience, facilitating deeper and more sustained learning outcomes.

Bouchée et al. [9] highlighted the need for students learning quantum physics to receive high-quality feedback from their teachers. This feedback makes students aware of misconceptions in their learning, guiding them towards more effective strategies for understanding quantum physics. In a related study, Abdurrahman et al. [47] investigated the impact of feedback in teaching quantum physics. Their findings indicated that feedback in quantum physics education significantly improved the scores of physics teacher candidates on the quantum physics concept inventory test. These studies emphasize the critical role of effective feedback in enhancing students' understanding and performance in quantum physics, supporting more accurate and deeper learning.

Some of the biggest challenges in giving effective feedback arise from teachers' limited time and heavy workloads or negative teacher or student attitudes [48, 49]. Giving students individual feedback, especially in crowded classroom environments, requires serious time and intense teacher performance, and is often impossible in practice [50, 51].

LLMs can analyse students' responses and leverage their theoretical awareness of effective feedback to provide comprehensive and varied comments [42], and they have potential in flipped classrooms [52]. Steinert et al. [42] developed a platform that uses the ChatGPT language model to provide formative feedback to students to enhance various self-regulated learning processes. On this platform, students answer questions provided by their teacher and may improve their answers after receiving personalized formative feedback from ChatGPT. Similarly, Uchiyama et al. [52] developed a platform where ChatGPT gives feedback on students' answers for questions related with an example situation via video; the researchers stated that this platform could eliminate the difficulties of giving feedback in flipped classrooms.

Jansen et al. [53] stated that the feedback given by LLMs was comparable in quality to the feedback given by experts; however, Steiss et al. [54] emphasized that human feedback was clearly superior compared to the feedback given by GPT 3.5. Jacobsen and Weber [55] argued that GPT-4 may surpass teachers in terms of general quality if trained with effective prompts. Dai et al. [56] found that ChatGPT can produce more detailed feedback that summarizes students' performance more fluently and consistently than human instructors, and, based on this, they stated that ChatGPT can provide feedback that will help students complete tasks and improve their learning skills.

Feedback can be given in different formats, such as written or verbal feedback. Agricola et al. [57] noted that verbal feedback had a significantly higher effect on students' perception of the feedback than written feedback, and a high perception of feedback is important for the feedback to be effective. Merry and Orsmond [58] stated that students found verbal feedback more positive and that verbal feedback was easier for them to understand because it was closer to dialogue. Van der Schaaf [59] gave written feedback to a group of 72 students, and then half of that group was given verbal feedback later in the study; they concluded that the students who received verbal and written feedback found it more effective than the students who did not receive verbal feedback.

Therefore, LLMs can be effectively integrated into personalized environments where they provide students with formative feedback. This allows them to overcome teachers' difficulty in giving personalized feedback to each student in a classroom environment. In addition, adding LLMs to AR environments and being able to conduct spoken dialogues with them allows them to effectively support students during experiments. Considering these potential benefits, LLMs' ability to be added to the AR environment to communicate verbally with students and give them feedback has a high potential to improve students' learning gains by facilitating personalized learning processes.

2.5 Eye-tracking

The eye–mind assumption states that gaze fixations and cognitive processing of the fixated content are closely linked [60, 61]. Accordingly, eye tracking is frequently used in learning sciences to investigate learning processes in detail [62–65]. Eye-tracking methods have already provided exciting insights into learning processes in physics education [66]. For example, during linking processes between the laboratory system and the coordinate system, a continuous animation that visualizes the translation process binds the gaze very strongly, whereas interactive measurement documentation tasks with feedback result in many cross-representational eye movements, which accompanied a more successful learning process [67]. Eye-tracking studies have also helped illuminate learning processes about vector fields [68, 69], graphs [70–73] or interactive learning systems in quantum physics [29]. In classical eye tracking on a 2D computer screen, research focused primarily on analysing the total number or duration of gaze fixations. In addition, the gaze duration in 'areas of interest,' i.e. specific sections of the 2D learning content, was often examined. The total number, direction, length and duration of saccadic movements, which are the eye movements connecting one fixation to the next, have also been examined [62, 74].

As this study did not involve a screen-based learning unit, innovative approaches must be chosen to investigate eye movements during real experimentation and interaction with virtual displays. To do so, the ARETT toolkit for the HoloLens 2 (HL2) was developed [75]. It allows volumes of interest to be defined. These volumes are registered as hit when the eye ray of the HL2 collides with the surface delimiting the volume. The toolkit has been used successfully to investigate eye movements in AR learning units and, especially, to analyse learning success [76].

For this study, ARETT was used to determine the total fixation duration on specific elements of the learning environment. In addition to demonstrating the learning effectiveness of AI-based feedback, this provides exciting insights into the gaze behaviours related to the learning processes that are triggered by feedback. Individuals' total fixation durations are associated with various cognitive processes. Studies indicate that total fixation

duration may be directly related to tasks' difficulty level. As the difficulty increases, individuals require more time and effort for information processing, which manifests as longer total fixation durations [77]. In addition, when an individual attempts to overcome a challenging task or encounters an unexpected impasse, an increase in total fixation duration is observed [78]. Moreover, an individual's level of experience significantly affects total fixation duration. Experienced individuals, having previously encountered similar tasks or situations, can process information more efficiently, which can result in shorter total fixation durations [79].

Finally, research reveals that total fixation duration is related to the intensity of cognitive processing [80] and, therefore, presents an informative indicator to interpret the effects of AI-based feedback.

3 Research questions

This study aims to investigate the effects of feedback provided by ChatGPT in AR-enhanced experiments in physics lab courses. To investigate the effects of integrating ChatGPT as a feedback tool, the following two research questions were analysed in a crossover design:

RQ1: Cognitive learning outcome: Does formative feedback via a large language model lead to a higher conceptual understanding during experimentation?

RQ2: Cognitive learning process: How does adding individualized ChatGPT-based feedback change students' visual attention while answering comprehension questions in a laboratory course on quantum cryptography?

4 Methods

4.1 Participants

In total, 21 groups of students (11 Group A, 10 Group B; sex: %63 male, %37 female; age: mean = 20.95; SD = 2.42), voluntarily participated in the experiment, which lasted about 3 hours.

All participants conducted the quantum cryptography analogy experiment in the laboratory following a lab manual specifically prepared for the experiment. All groups came to the laboratory at different times during the lab course.

4.2 Experimental setup

The quantum cryptography analogy experiment set developed by Thorlabs is based on the BB84 protocol. It includes an experimental setup using short-duration laser pulses instead of single photons. The experiment set primarily consists of a laser source, lambda/2 plates and detectors. Students adjust the angles of the lambda/2 plates to determine the bases and bits of the emission and use the detectors to measure the incoming bits.

In the AR-enhanced version of the experiment, vectorial representations are used to visualize the polarization state of a single photon. QR codes are attached to the experimental setup to ensure that the AR glasses display the visualizations in the correct locations of the setup. As soon as the glasses recognize the QR codes, the corresponding visualizations are displayed. A microcontroller reads out the states of the lambda/2 plates in the setup in real time and sends this data to the AR glasses via Bluetooth. This enables the current polarization states in AR to be displayed in real time according to the lab partners' actions.

The AR application incorporates the most immersive features of AR systems, such as combining real and virtual objects in a real environment, operating interactively and in real-time and aligning real and virtual objects with each other [32]. The 3D visuals on

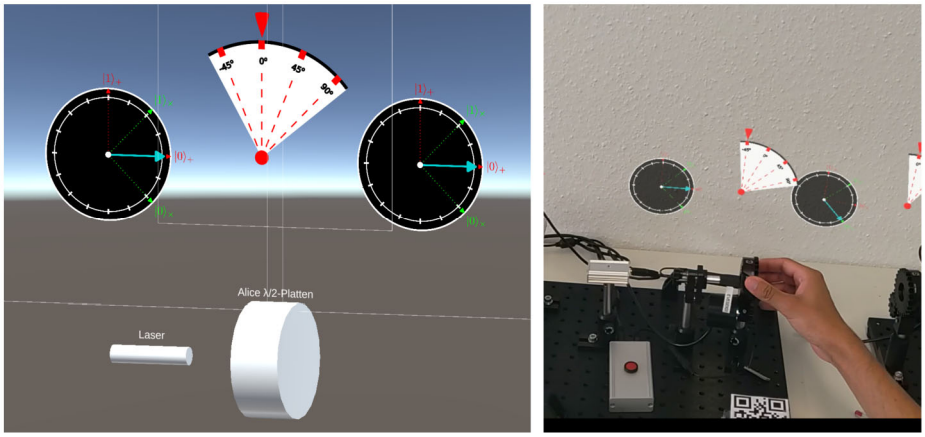

Figure 1 AR visualization in the application. The elements with a white background show the angle of the lambda/2 plate and update in real time if the physical setup is changed. In the circular elements with a black background, the polarization state of a single photon is represented with vector notation; these also update according to changes in the angle of the real lambda/2 plate

the right sides of each lambda/2 plate in the real world update dynamically according to the lambda/2 plate's angle. These visualizations allow students to observe the polarization state before and after it reaches the lambda/2 plate, as well as the effect of the angle of the lambda/2 plate on polarization. The AR visualizations included in the application are shown in Fig. 1. Elements with a white background show the angles of the lambda/2 plates, and elements with a black background show the photons' polarization state. Both visualizations update according to the angles of the lambda/2 plates.

4.3 Objectives in the laboratory course

The primary objective of the quantum cryptography analogy experiment is to exchange bits using the BB84 protocol to generate a key for secure communication and to test, through different methods, whether an agent named Eve was eavesdropping on the exchange. The students followed the protocol twice—once without Eve in the system and once with Eve present. Their task was to determine the key code and test for Eve's presence. In addition, at four different times during the experiment, ChatGPT verbally asked the students a question related to the previously completed step.

All steps of the process are detailed in Table 2.

Table 2 Experimental process steps

Step	Task
1	Base-bit determination (for Alice); base determination (for Bob)
2	Sending and receiving bits (Alice and Bob)
3	Base comparison, key determination
4	First comprehension question
5	Encrypting, transmitting and decrypting a secret message
6	Second comprehension question
7	Base-bit determination (For Alice); base determination (for Bob and Eve)
8	Sending and receiving bits (Alice, Bob and Eve)
9	Third comprehension question
10	Base comparison, key determination, selection of a random bit from the key and comparison
11	Fourth comprehension question

After listening to the questions in Tasks 4, 6, 9 and 11, the students determined their answers and communicated them to ChatGPT. Then, depending on their assigned group, the students either received feedback from ChatGPT or did not (see Sect. 4.5).

4.4 ChatGPT usage in the application

During the lab course, students must perform different lab tasks. For specific tasks (see Sect. 4.3) the students interact with ChatGPT by asking questions and receiving answers. To verbally interact with ChatGPT, the speech-to-text feature first converts the students' spoken words into text, which is then sent to ChatGPT as an input. Subsequently, the written output from ChatGPT is converted into speech with a text-to-speech function, allowing students to hear the response (see Additional file 1). The ChatGPT menu in the application is shown in Fig. 2.

To provide feedback, the OpenAI API and the GPT-4 model [81] were used. For ChatGPT to give feedback on students' verbally formulated answers, pre-prompts were sent along with the request.

i. The prompts prepared for questions with feedback include:

- The relevant question and its answer.
- Information that the prompt comes from an AR-assisted experiment conducted by students.
- A command to ChatGPT to ask the question without mentioning that it is an AI bot.
- A directive that when ChatGPT receives an answer from students, it should act like a teacher and compare the answer with the correct answer. In cases of incorrect or incomplete responses, it should guide students toward the correct answer without providing information about the correct answer.
- Tell students to proceed to the next stage once their answer is considered correct.

ii. The prompts prepared for questions without feedback include:

- Only the relevant question.
- Information that the prompt comes from an AR-assisted experiment conducted by students.
- A command to ChatGPT to present the task without mentioning that it is an AI bot.

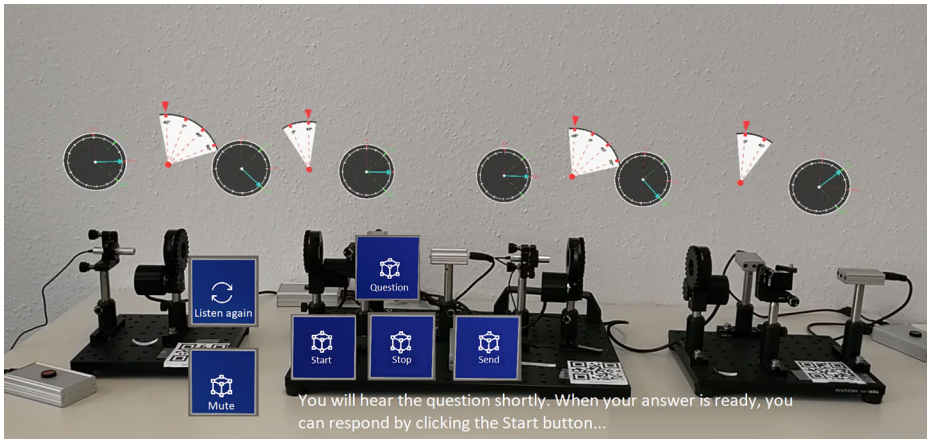

Figure 2 Interface for interaction with ChatGPT

- A directive to tell students to proceed to the next stage without providing feedback after receiving their response.

4.5 Study design

At four points in the lab course, students were asked to answer comprehension questions that referred to previous tasks. The students responded verbally to these questions. The use of ChatGPT for these questions involved two distinct methodologies: with feedback (F) and without feedback (W). In the feedback sections, ChatGPT provided feedback by comparing students' responses with the correct answers. Feedback guides students to improve their answer accuracy through discussion in cases of incomplete or incorrect responses. Students were randomly divided into two groups, A and B. Group A received ChatGPT-based feedback on their answers to Questions 1 and 3 and Group B received ChatGPT-based feedback on their answers to Questions 2 and 4. Although only the student interacting with ChatGPT receives feedback, the feedback is delivered through the HL2 speaker, so the entire lab course group benefits. Students can therefore work together in the group to find the best possible answer to the individual tasks using the feedback. The research design is shown in Fig. 3.

4.6 Instruments

4.6.1 Cognitive learning outcome: questions about experimental processes

The first research question was about the impact of ChatGPT feedback on students' performance scores when answering questions. Each question's answer contained four parts, each part assigned 0.25 points, and a complete explanation of each question was assigned 1 point. These questions were integrated into the prompts sent to ChatGPT, and the students' conversations with ChatGPT were documented in text form. The first question in-

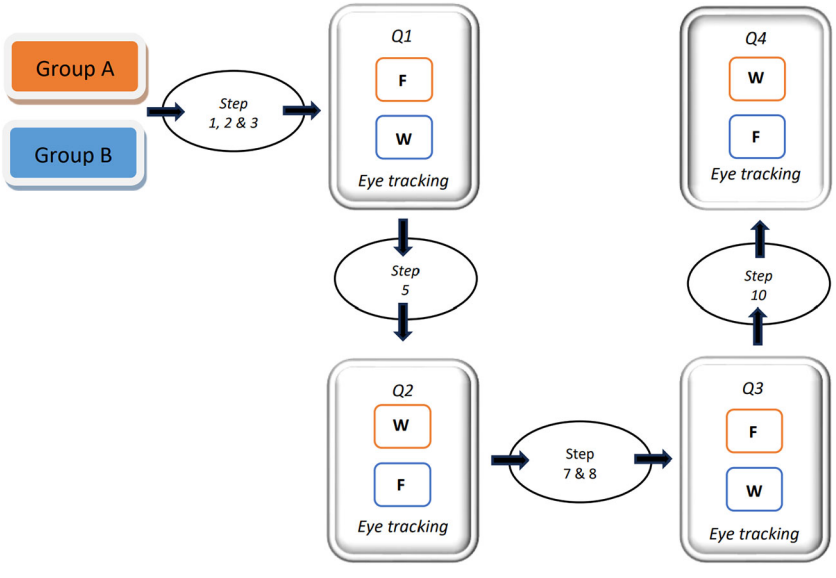

Figure 3 Research design of the study. Both groups performed the same experimental steps (see Table 2). After completing specific steps, the groups answered Questions 1, 2, 3 and 4 (Q1, Q2, Q3 and Q4), either with feedback (F) or without feedback (W). During the Q/A stages, eye-tracking data were collected from the students

volved explaining why, in quantum cryptography, two bases are used. The second question asked students to name the necessary aspects to generate a key. The third question involved describing the experimental procedure that Eve uses to try to intercept the communication, and the fourth question asked how it is possible to discover whether a spy has intercepted a communication by comparing bits and bases. Overall, questions 1, 2 and 4 refer to more conceptual knowledge, while question 3 refers to an experimental approach.

4.6.2 Cognitive learning process: eye gaze data collection with ARETT

To answer the second research question, eye gaze data were collected with the ARETT toolkit to analyse the students' total fixation duration while responding to ChatGPT feedback. The ARETT toolkit is specially developed to easily collect eye gaze data in AR applications. ARETT consists of a data access layer that makes timestamped gaze data available in real time, a data recorder to store the data, a web-based control interface and a set of auxiliary tools for data visualization. By integrating this toolkit into their applications, developers can collect data in a controlled manner through the web-based interface. Detailed information about ARETT can be found in the study of Kapp et al. [75].

After adding the ARETT toolkit to the application, specific areas of interest (AOIs) were created for the physical and virtual components of the experiment. AOIs are used in eye-tracking studies to define specific regions where participants are likely to focus their attention. Physical AOIs are connected to the components at the experimental set and the virtual AOIs are connected to AR representations.

4.7 Data analysis

4.7.1 Cognitive learning outcome: data analysis on the effects of ChatGPT feedback

The file containing written transcripts of students' verbal conversations with ChatGPT throughout the experiment was transferred from HL2 to a computer at the end of the experiment. Important aspects that are necessary to answer the question correctly were determined in advance. The scoring was then carried out on the basis of these considerations.

To test for the presence of a carry-over effect, the sum score of the four questions was calculated for both Group A and Group B. Then, these scores were tested for differences between the groups. The carry-over effect is examined to determine whether the different experimental conditions influence each other [82]. In this study, the potential carry-over effect was examined to see whether receiving and not receiving feedback influenced students' work (for example, if the feedback students receive on Q1 affects their answer to Q2). This effect describes the persistence of the impact of one method (receiving feedback on responses) on the subsequent method (not receiving feedback on responses). In crossover design studies, the presence of a carry-over effect can jeopardize the reliability and validity of the results [83]. Therefore, controlling for this effect is essential.

To compare the scores of questions with feedback to those without feedback, the differences (D) between the scores for each group were initially calculated with the following equation:

$$D = [(Q1 + Q3) - (Q2 + Q4)] \tag{5}$$

Q1 and Q3 are the scores of the questions for which Group A received feedback, while Q2 and Q4 are the scores of the questions for which Group B received feedback. If the

scores are high when feedback is received, Group A's D-scores will be higher than Group B's D-scores. As the difference between the scores with and without feedback increases, the difference between the mean D-scores of the two groups will also increase. This allows scores in the feedback and non-feedback conditions to be compared using the D-scores. The D-scores for Groups A and B were compared with independent sample t-test.

4.7.2 Cognitive learning process: data analysis for students' total fixation durations during ChatGPT interactions

The eye-tracking data collected with ARETT was automatically saved on the HL2 as a .csv file when the data collection process was completed.

Total fixation duration metrics were used to analyse the collected eye-tracking data. Total fixation duration gives information about visual attention on the predefined AOIs related to physical components and virtual representations while answering the questions. Fixation duration measures the specific time a student spends looking at a particular AOI. Throughout the process, all such gazes are aggregated to calculate the total fixation duration. During students' responses to questions, the total fixation durations were first calculated for each AOI and each participant. Then, these totals for each participant were divided by that participant's total fixation duration for all AOIs to obtain the normalized total fixation duration for specific components of the experiment. The subsequent analyses were conducted using these normalized total fixation durations. This method determined how visual attention was distributed between the two AOIs.

A paired-sample t-test was used for the quantitative analyses related to the gaze data. Since these analyses were conducted only on students who received feedback, the sample size is limited. To minimize the risk of overlooking significant differences, the significance level for the p-value was set at 0.10. This choice is based on the requirement for a larger sample size to reliably detect any effects, as noted in the study by Cronbach and Snow [84]. Since the recommended sample size from the power analysis was not fully met, the alpha error level was increased to optimize the beta error and reduce the likelihood of missing interaction effects.

5 Results

In this section, the results will be presented. First, the results regarding the effects of ChatGPT feedback on students' answers will be described. Then, the results of the analyses of the normalized total fixation duration will be presented.

5.1 Learning outcome results related to the effects of ChatGPT feedback

During the answering process, no limitation was imposed on the amount of feedback students could receive. They could continue receiving feedback on their responses until ChatGPT was convinced that their answer was sufficient. Receiving multiple rounds of feedback indicates that the students' responses were inadequate or incorrect and that they needed more feedback to improve their answers. The number of feedback responses received by the groups is presented in Table 3.

Table 3 shows that a high percentage of participants responded to Q1 and Q4 without receiving feedback. Most participants received feedback on Q2 and Q3 at least once.

The average scores for each question and the impact of feedback on these scores are presented in Table 4 and the bar graph in Fig. 4.

Table 3 Number of groups broken down by the amount of feedback received

	Q1 (Group A)	Q2 (Group B)	Q3 (Group A)	Q4 (Group B)
0 feedback	6	4	4	7
1 feedback	5	2	5	2
2 feedback		3	2	1
3 feedback		1		

Table 4 The average score of learning outcome (X) and standard error (SE) of the points groups received from answering questions. The 'before feedback' values are based on the points students would have received from their initial answers before any feedback was provided. The 'after feedback' values were calculated using the total points from all answers given to the question (for feedback groups)

	Group	Before feedback		After feedback	
		$\overline{X}$	SE	$\overline{X}$	SE
Q1	A	0.454	0.106	0.500	0.095
	B	0.400	0.100	–	
Q2	A	0.523	0.132	–	
	B	0.350	0.135	0.675	0.084
Q3	A	0.295	0.030	0.636	0.070
	B	0.200	0.033	–	
Q4	A	0.477	0.071	–	
	B	0.350	0.100	0.500	0.092

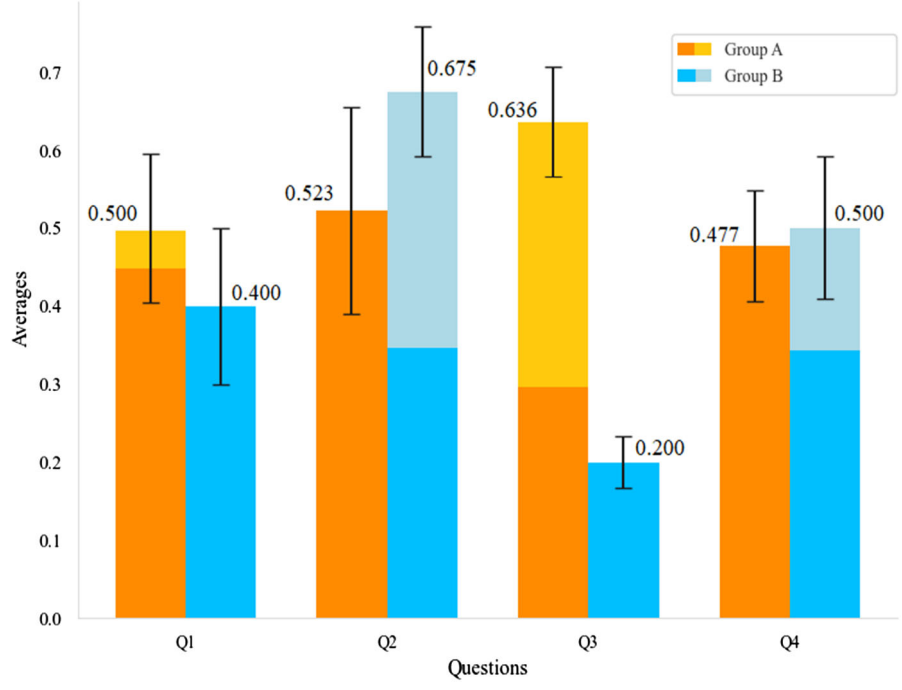

Figure 4 The average scores for each question and the change in score when feedback was received. Dark bars show the scores of the initial responses; light bars show the average increase in score after receiving feedback; the error bars represent the standard error for the total points, and the numerical values of these data are presented in Table 4

Table 5 Independent sample *t*-test results used to analyse the difference between the groups' *D*-scores, which represent the difference between their response scores with and without feedback

Group	*N*	$\overline{X}$	SD	*t*	*p*	Cohen's d
A	11	0.136	0.595	2.721	0.014*	1.189
B	10	–0.575	0.602			

*indicates a significant ($p < 0.05$) difference.

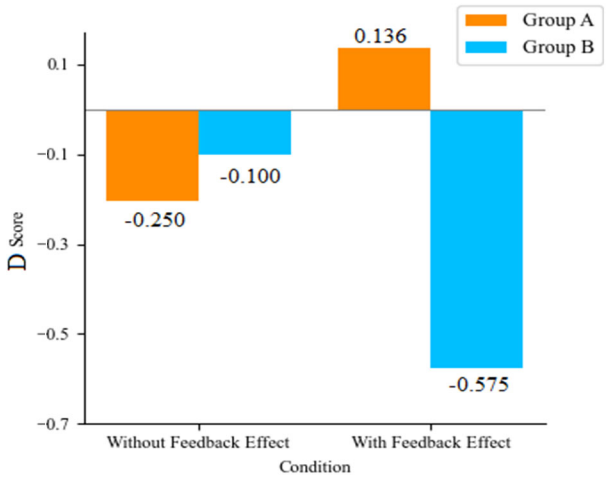

Figure 5 *D*-score averages of the groups. The bars on the left are the *D*-scores calculated using only the scores students received from their initial answers, without feedback. The bars on the right are the *D*-scores calculated based on the total scores obtained with feedback

To conduct the statistical analysis, first, the total scores of the groups on all four questions (Q1 + Q2 + Q3 + Q4) were compared to check for any carry-over effect. As the total scores for both groups showed a normal distribution, an independent sample *t*-test was used to compare the scores. The *p*-value was determined as 0.265 in this test. Therefore, the difference between the total scores of the two groups was not statistically significant ($p > 0.05$). This indicates that there was no carry-over effect.

After ensuring that no carry-over effect occurred, the *D*-scores (Q1 + Q3 − Q2 − Q4) for both groups were calculated with Eq. (5). For the comparison of the *D*-scores between the two groups, independent sample *t*-test was used. The results obtained are presented in Table 5.

As Table 5 shows, the difference between the *D*-scores for the two groups is statistically significant with a large effect ($t = 2.721$, $p = 0.014$, $d = 1.189$). The *D*-scores of Group A are higher than those of Group B; therefore, the significant difference is in favour of the questions with feedback.

To conduct a more detailed analysis of whether the observed differences in scores were due to feedback, *D*-scores were calculated for both groups' scores without feedback. In this process, the scores from the students' initial responses to the questions were considered, while the scores obtained after receiving feedback were ignored. The new *D*-scores were then analysed with the independent sample *t*-test, which resulted in a *p*-value of 0.700. Therefore, when the effect of feedback is disregarded, the *D*-scores of the groups do not show a statistically significant difference.

The average *D*-scores of the two groups without and with feedback are presented in Fig. 5.

As Fig. 5 shows, the groups' D-scores are quite close to each other when the feedback effect is omitted and quite divergent when the feedback effect is included. When the effect of feedback is incorporated into the scores, the average score for Group A is 0.136, while the average score for Group B is −0.575. Group B's lower scores compared to Group A indicate that the scores for the questions with feedback are higher than those for the questions without feedback.

5.2 Results of cognitive learning process expressed by students' normalized total fixation duration during ChatGPT interactions

To achieve a more detailed analysis of the impact of ChatGPT feedback, each ChatGPT-supported question's initial score and score after receiving feedback were compared. The results of the paired sample t-test for this comparison are presented in Table 6.

Examining Table 6 reveals significant differences in favour of the post-feedback condition for Questions 2, 3 and 4. To further explore the impact of ChatGPT feedback on the response process that contributed to these differences, an eye gaze analysis was conducted. This analysis examined students' visual attention before and after receiving feedback, specifically during the first feedback session, as the goal was to directly compare the data without feedback and with feedback. The second and third feedback sessions are likely to carry over the effects of the first feedback session.

In this section, the results concerning the normalized total fixation durations during the response processes for Q1, Q2, Q3 and Q4, following the research design flow outlined in Fig. 3, will be discussed in detail. The main aim of the second research question was to analyse the effect of ChatGPT feedback on visual attention. To achieve this objective, the students' visual attention to the physical AOIs of the physical experiment set and the virtual AOIs of the AR environment were compared. The effect of feedback on the differences in attention paid to these two types of AOIs was also analysed. To achieve realistic results for this research question, only the eye gaze data of the group members who used the ChatGPT function were used in the analysis.

For the first question, five groups received feedback (see Table 3). Figure 6 illustrates the distribution of the normalized total fixation duration for these groups before and after receiving feedback.

A paired samples t-test was performed to evaluate whether a difference exists between the normalized total fixation durations on physical and virtual components. The results are presented in Table 7.

Table 6 Paired sample t-test results comparing scores for each question with and without ChatGPT feedback

		N	$\overline{X}$	SD	t	p	Cohen's d
Q1	W	11	0.454	0.350	−1.000	0.170	–
	F	11	0.500	0.362			
Q2	W	10	0.350	0.428	−3.545	0.003*	0.913
	F	10	0,675	0.265			
Q3	W	11	0.295	0.101	−4.892	<0.001*	1.892
	F	11	0.636	0.234			
Q4	W	10	0.350	0.316	−1.964	0.040*	0.495
	F	10	0.500	0.289			

*indicates a significant ($p < 0.05$) difference; (Q: Question, W: Without feedback, F: With feedback).

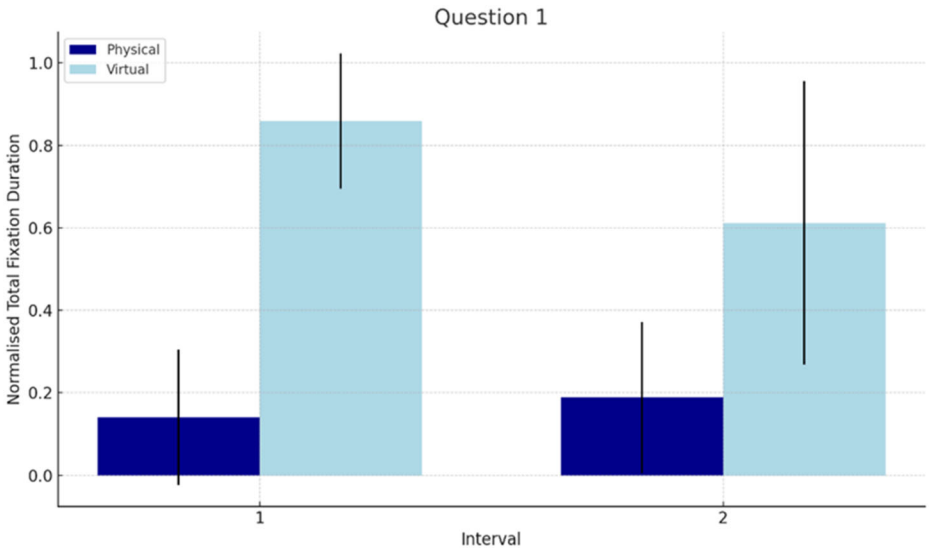

Figure 6 Plots of the average normalized total fixation duration when answering the first question. Interval 1 represents the pre-feedback period, while Interval 2 represents the post-feedback period. The bars indicate the average normalized total fixation duration and the standard deviation

Table 7 Paired sample *t*-test results comparing the normalized total fixation durations on physical and virtual AOIs while answering the first question

	AOI	*N*	$\overline{X}$	SD	*t*	*p*	Cohen's d
Interval 1	Physical	5	0.141	0.184	−4.372	0.012*	3.902
	Virtual	5	0.859	0.184			
Interval 2	Physical	5	0.188	0.204	−2.237	0.089*	1.379
	Virtual	5	0.612	0.384			

*indicates a significant ($p < 0.10$) difference.

The results indicate that the normalized total fixation duration on virtual components is significantly higher than that on physical components during both intervals.

For the second question, six groups received ChatGPT feedback. Figure 7 shows the distribution of their normalized total fixation durations on physical and virtual components during Interval 1 and Interval 2.

A paired samples *t*-test was performed to evaluate whether a difference exists between the normalized total fixation durations on physical and virtual components during the two intervals. The results are presented in Table 8.

The results indicate that the normalized total fixation duration on virtual components is significantly higher than that on physical components during both intervals.

For the third question, one group was excluded from the analysis due to technical problems during the intervention. Eye gaze analyses for the third question were performed specifically for the six groups who received feedback. Figure 8 illustrates the distribution of their normalized total fixation durations on physical and virtual components during Interval 1 and Interval 2.

A paired samples *t*-test was performed to evaluate whether a difference exists between the normalized total fixation durations on physical and virtual components during the two intervals. Table 9 presents the results.

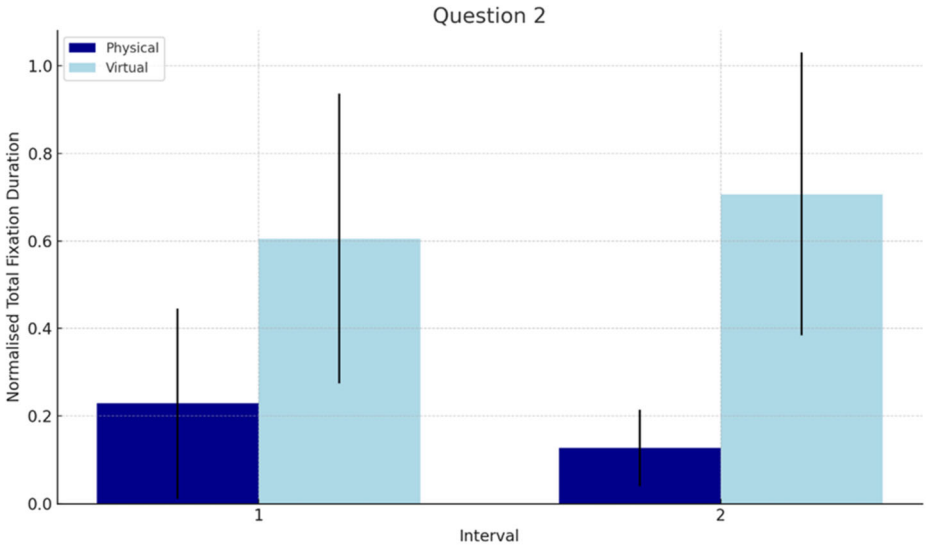

Figure 7 Plots of the average normalized total fixation duration when answering Q2. Interval 1 represents the pre-feedback period, while Interval 2 represents the post-feedback period. The bars indicate the average normalized total fixation duration and the standard deviation

Table 8 Paired sample *t*-test results comparing the normalized total fixation durations on physical and virtual AOIs while answering the second question

	AOI	*N*	$\overline{X}$	SD	*t*	*p*	Cohen's *d*
Interval 1	Physical	6	0.228	0.238	–2.014	0.099*	1.228
	Virtual	6	0.605	0.363			
Interval 2	Physical	6	0.127	0.095	–4.460	0.007*	2.238
	Virtual	6	0.707	0.354			

*indicates a significant ($p < 0.10$) difference.

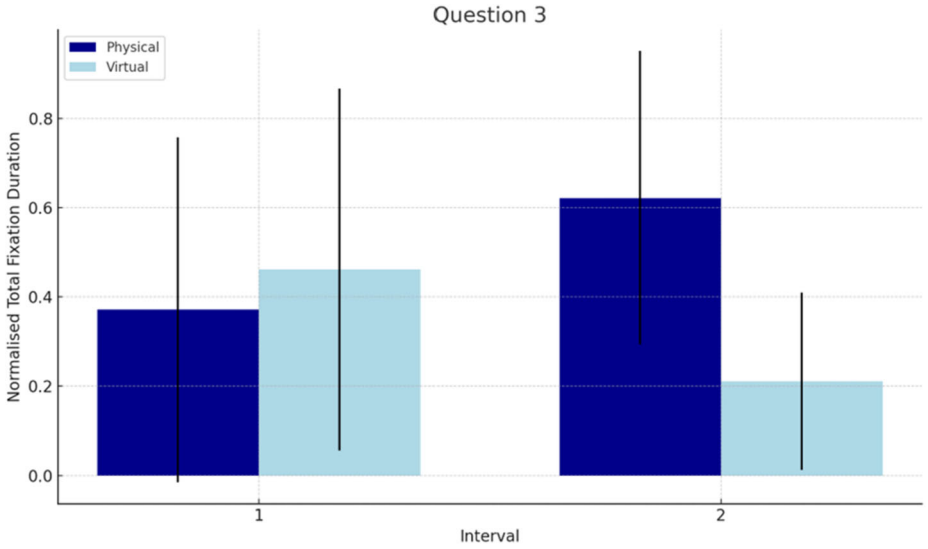

Figure 8 Plots of the average normalized total fixation duration when answering the third question. Interval 1 represents the pre-feedback period, while Interval 2 represents the post-feedback period. The bars indicate the average normalized total fixation durations and the standard deviation

Table 9 The paired sample *t*-test results comparing the normalized total fixation durations on the physical and virtual AOIs while answering the third question

	AOI	*N*	$\overline{X}$	SD	*t*	*p*	Cohen's *d*
Interval 1	Physical	6	0.372	0.423	−0.285	0.785	–
	Virtual	6	0.462	0.444			
Interval 2	Physical	6	0.622	0.360	2.233	0.068*	1.379
	Virtual	6	0.211	0.218			

*indicates a significant ($p < 0.10$) difference.

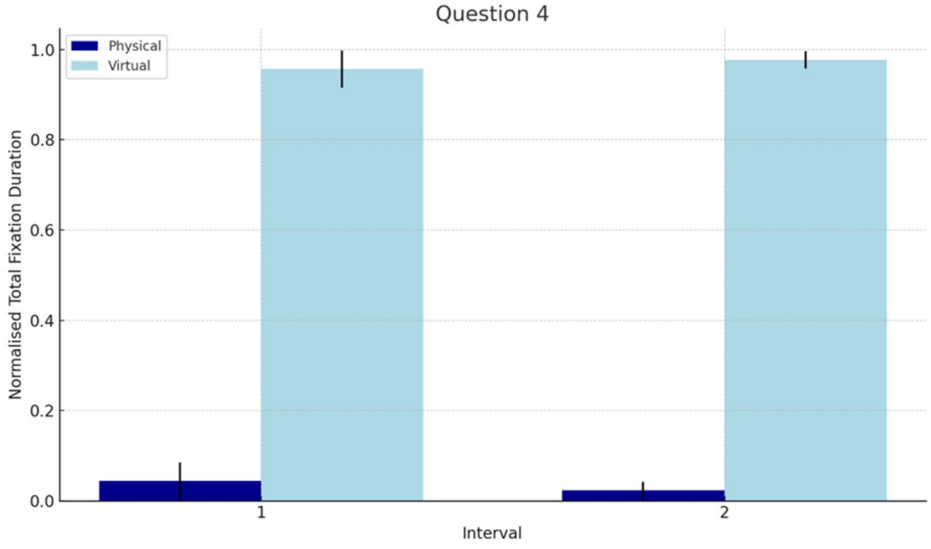

Figure 9 Plots of the average normalized total fixation duration while answering the fourth question. Interval 1 represents the pre-feedback period, while Interval 2 represents the post-feedback period. The bars indicate the average normalized total fixation durations and the standard deviation

Table 10 Paired sample *t*-test results comparing the normalized total fixation durations on the physical and virtual AOIs while answering the fourth question

	AOI	*N*	$\overline{X}$	SD	*t*	*p*	Cohen's *d*
Interval 1	Physical	3	0.044	0.050	−15.813	0.004*	18.240
	Virtual	3	0.956	0.050			
Interval 2	Physical	3	0.023	0.023	−35.774	0.001*	41.478
	Virtual	3	0.977	0.023			

*indicates a significant ($p < 0.10$) difference.

The *t*-test results indicated that no significant differences exist between the normalized total fixation duration on virtual and physical components during Interval 1. However, after ChatGPT feedback, during Interval 2, the normalized total fixation duration on physical components is significantly higher than that on virtual components.

A total of three groups received feedback on Q4. Figure 9 illustrates the distribution of the groups' normalized total fixation durations on physical and virtual components during Interval 1 and Interval 2.

A paired samples *t*-test was performed to evaluate whether a difference exists between the normalized total fixation durations on physical and virtual components during the two intervals. Table 10 presents the results, which indicate a significant difference in the

distribution of gaze during both intervals. The virtual elements received significantly more visual attention than the physical elements.

The results indicated that the normalized total fixation duration on virtual components is significantly higher than on the physical components for both intervals.

6 Discussion

6.1 Discussion of the learning outcomes related to ChatGPT feedback

To answer the first research question, the differences in scores on questions with and without feedback were compared. The results showed that ChatGPT feedback had a significant positive impact on students' scores. The average scores on questions with feedback are consistently higher than those on questions without feedback.

Feedback increases students' scores on each question. The biggest difference between the feedback group and the without-feedback group was observed on Question 3, which asked about the principle Eve used to obtain information. The second highest difference in scores occurred on Question 2, which asks about the features an encryption key should have. The difference in scores was lower on Question 1, which asks why two bases are used. On Question 4, which asks about the method Alice and Bob follow to determine whether an intercepting agent is present, the average scores of the two groups exhibit almost no difference (see Table 4 and Fig. 4).

Most groups answered the first and fourth questions without receiving feedback (see Table 3), and ChatGPT's feedback had a minimal effect on their scores compared to the second and third questions (see Fig. 4). During the application's programming phase, the responses to the first and fourth questions in the prompts sent to ChatGPT were provided in a single paragraph without clearly distinguishing the sub-sections. This likely led ChatGPT to interpret incomplete answers as correct. Consequently, students moved on without receiving any feedback. In contrast, for the second and third questions, the sub-parts of the answers were clearly separated by bullet points, resulting in more students receiving feedback and, subsequently, achieving higher scores. This situation maybe highlights the importance of clearly distinguishing the content within responses provided in prompts used during the programming phase to ensure effective use of ChatGPT.

One of the key points to highlight in this study is that ChatGPT was not used to provide information but rather as a tool to guide students towards discovering information. The ChatGPT feedback did not contain any direct information related to the correct answers. To ensure this, the prompts sent to ChatGPT were meticulously prepared to ensure that it avoided offering any answer-related information. Instead, the feedback was designed to encourage students to think more deeply and foster group discussions about missing points. Students used ChatGPT as a virtual group partner, and the feedback they received drew them to focus more on the points they had probably missed in their pre-feedback responses. During the post-feedback intervals, they engaged in more intense and small-scale discussions. As noted in the literature [24], these small discussions increased student success. In addition, as the literature suggests, students improved their learning outcomes in this study through ChatGPT-supported small discussions.

ChatGPT's ability to provide instant feedback has been the subject of numerous studies [53–56]. In parallel with these, the findings of this study demonstrate that ChatGPT feedback improved students' responses to the four questions asked throughout the quantum cryptography analogy experiment.

6.2 Discussion of the learning process related to the normalized total fixation duration

On the first, second and fourth questions, visual attention to virtual components was significantly higher during both intervals. These questions were more theoretical and conceptual in nature. Research suggests that experimenting in virtual environments supports conceptual understanding more effectively than real-world experimentation [85–88]. Therefore, in an experimental process involving both physical and virtual components, focusing more on virtual representations while answering conceptual questions may facilitate a correct answers. The observation that students paid significantly greater attention to virtual representations when answering these concept-based questions, resulting in improved scores, supports this conclusion. In addition, ChatGPT feedback played an important role in keeping students' attention on the virtual representations, ensuring that they remained focused after receiving feedback.

In the third question, no statistical differences were observed during the first interval, but in the second interval, after receiving feedback, a significant increase in visual attention to the physical components was observed. When students initially encountered the question, they showed similar levels of attention to both types of AOIs. However, after receiving feedback, they shifted their attention to the physical components. The third question asked about Eve's behaviour during the experimental process, and the knowledge assessed by this question was technical and related to the experiment. Research shows that hands-on activities are significantly more effective for developing technical knowledge about experimental tools [85–87, 89]. Therefore, increased attention to the physical experimental setup is expected to lead to more accurate answers to technical questions. Consequently, the shift in visual focus after receiving ChatGPT feedback can be seen as supporting correct answers to this question. The change in attention between intervals demonstrates how effective ChatGPT feedback was in guiding students' attention to improve the accuracy of their responses.

6.3 Limitations

This study was limited to 21 groups of students. Participation was entirely voluntary. While this approach ensured that the findings were based on genuine, unbiased experiences, it also means that the sample may not fully represent the broader student population. As participants engaged in the study out of personal interest and curiosity, selection bias may have occurred.

The conversations conducted with ChatGPT were confined to the state of GPT-4 between March 6th and May 23rd, 2024. Therefore, the accuracy of the feedback provided by ChatGPT is limited by the quantum cryptography and BB84 protocol knowledge of the version available during this period.

None of the participating students reported prior experience using HL2. Many students initially struggled to use the buttons within the application. However, they quickly adapted and continued to use it without issue during the subsequent sections. In addition to their lack of experience, most students mentioned that towards the end of the experimental process, the weight of the HL2 glasses and the heat from the battery started to cause some discomfort.

The study compares a feedback condition with a non-feedback condition. Normally, one would expect that giving feedback automatically leads to better results, but this was

not self-evident for AI-generated feedback in such a complicated subject area as quantum physics and should be verified in this study. The study shows that ChatGPT is able to correctly interpret answers to very complex physical questions and provide adaptive, individualized feedback, which also led to an improvement in the answer. In the future, further studies should examine the extent to which other forms of feedback, such as giving the correct answer, compare to feedback generated by ChatGPT.

7 Conclusions

The first research question of this study considered the impact of ChatGPT feedback on student responses. The findings demonstrated a significantly higher total score if the students received ChatGPT feedback. This suggests that ChatGPT feedback improved students' learning outcomes and the quality of their responses. These results support the integration of AI-driven feedback systems in educational AR settings by demonstrating that such tools can significantly improve learning outcomes.

The second research question of this study investigated the effects of ChatGPT feedback on students' gaze behaviour. ChatGPT feedback tailored to the students' responses had varying effects on their visual attention. This is an important finding in the context of personalized learning environments. For some questions, ChatGPT feedback effectively maintained students' existing visual focus. However, for one question, ChatGPT feedback shifted students' visual attention, resulting in a significant difference in attention to physical and virtual AOIs. Considering the question content and the expected gaze behaviours, ChatGPT feedback produced the desired effect on all four questions. It kept students' attention on virtual components for conceptual questions, while for the question related to the experimental process, it shifted their focus towards the physical AOIs. This indicates that ChatGPT can direct students' visual attention via personalized feedback during the question-answering process.

It is important to note that students initially show a clear tendency to focus their visual attention to the virtual elements to answer the questions. This may be because in an AR-based learning environment, the new and innovative representations are particularly eye-catching for learners. However, the ChatGPT-based feedback leads to the focus of visual attention being placed on the real experimental equipment in more process-oriented activities such as question 3. The opposite is observed for more concept-related questions such as question 2. After receiving feedback, students focus more on the virtual elements. Therefore, ChatGPT-based feedback seems to be particularly helpful for AR-based learning environments as it specifically focuses visual attention on the necessary elements depending on the question. This in turn helps the learner to concentrate on the essentials.

The results of RQ2 show that ChatGPT feedback effectively manages students' visual attention during the learning process. Similarly, the findings of RQ1 indicate that these processes significantly contribute to the improvement of students' learning outcome. In addition to evaluating these two results independently, it is important to establish a connection by considering the processes and the outcomes in relation to each other. The parallel alignment between the learning process data and learning outcomes is a particularly delighted finding. The simultaneous examination of these two results demonstrates that students who receive feedback from ChatGPT increase their learning outcomes by engaging in effective cognitive learning processes while searching for the correct answer. They shifted their gaze to elements which are more important to answer a specific question or

kept their gaze focused on important elements while thinking about the correct answer. The alignment between the results of the two research questions supports the conclusion that ChatGPT's approach of comparing students' responses with the correct answers and providing personalized feedback may be used effectively in AR-based lab courses in quantum physics education.

The study was conducted to test the general feasibility and possibilities of the learning approach. Since no study has ever been conducted before using the combination of head-mounted-AR, AI support and eye tracking to support and investigate complex learning in a laboratory course, we deliberately decided to use a smaller sample size, as this is initially sufficient to check the feasibility of the laboratory course and to provide initial insights into the effects on the learning outcomes and the learning process.

However because of the low level sample size, these results are limited. As the reported results are very promising, in future this approach will be analysed more comprehensively using a larger sample size. This will allow for a more detailed examination of how ChatGPT feedback affects learning in AR environments.

Educational processes must closely follow and effectively utilize emerging technologies to train individuals who will contribute to the further development of these technologies. Although this field has a limited number of studies, they indicate important effects due to mixed reality and artificial intelligence integration [90–93]. The quantitative results obtained in this study similarly demonstrate that the use of LLMs in AR environments improves student performance outcomes. Furthermore, findings related to student gaze data indicate that LLM feedback plays a crucial role in personalized learning and guiding students' focus towards more important experimental components for answering questions. Future research focusing on various uses of LLMs in AR/VR environments and their effects on different dependent variables would deepen the theoretical understanding. Furthermore, eye gaze analyses in these environments are suitable to gain more comprehensive insights and draw inferences regarding the effects of LLMs on student gaze patterns.

8 Controlled variables

The system usability scale (SUS) was administered at the end of the experiment to assess students' perceptions of the usability of the experimental tools. An analysis of the results revealed an average score of 76.316 (SD: 10.295). This finding indicates that students rated the usability of the application as 'Excellent' [94], suggesting that there was no usability-related limitation of the study.

Abbreviations

AR, Augmented reality; AI, Artificial intelligence; LLMs, Large language models; HL2, HoloLens 2; AOIs, Area of interests; MERs, Multiple external representations; VR, Virtual reality.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjqt/s40507-025-00310-z.

Additional file 1. (MP4 180.6 MB)

Author contributions

All authors conceptualized the research.

Funding
Open Access funding enabled and organized by Projekt DEAL. This research was funded by the German Research Foundation(Deutsche Forschungsgemeinschaft, DFG; Project "Real-world anchored augmented visualizations to support conceptual learning in physics"; Project number: 517713394) and supported by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

Data availability
The supporting data for this study is not publicly available. However, upon reasonable request, the corresponding author can provide access to the data for verification and replication purposes. Requests for data access should be directed to Atakan Coban at atakan.coban@lmu.de.

Declarations

Ethics approval and consent to participate
The data was collected following the approval of the responsible ethics committee and with the consent of the students.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Chair of Physics Education, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU), Munich, Germany.
[2]Physics Laboratory Courses, Faculty of Physics, Ludwig-Maximilians-Universität München (LMU), Munich, Germany.

Received: 30 October 2024 Accepted: 7 January 2025 Published online: 29 January 2025

References

1. Krijtenburg-Lewerissa K. Teaching quantum mechanics at the secondary school level. 2020. https://doi.org/10.3990/1.9789036549905.
2. Gyongyosi L, Imre S. A survey on quantum computing technology. Comput Sci Rev. 2019;31:51–71. https://doi.org/10.1016/j.cosrev.2018.11.002.
3. Wang J, Sciarrino F, Laing A, Thompson MG. Integrated photonic quantum technologies. Nat Photonics. 2020;14(5):273–84. https://doi.org/10.1038/s41566-019-0532-1.
4. Pirandola S, Andersen UL, Banchi L, et al. Advances in quantum cryptography. Adv Opt Photonics. 2020;12(4):1012–236.
5. Faerman V, Terekhin AO, Bragin D, Shelupanov A. Design and modeling of hardware kit for QKD education of engineering students and communication engineers. 2024. https://doi.org/10.20944/preprints202401.1754.v1.
6. Kaur M, Venegas-Gomez A. Defining the quantum workforce landscape: a review of global quantum education initiatives. Opt Eng. 2022;61:1–33. https://doi.org/10.1117/1.oe.61.8.081806.
7. Fox MF, Zwickl BM, Lewandowski HJ. Preparing for the quantum revolution: what is the role of higher education? Phys Rev Phys Educ Res. 2020;16(2):020131. https://doi.org/10.1103/PhysRevPhysEducRes.16.020131.
8. Bitzenbauer P. Quantum physics education research over the last two decades: a bibliometric analysis. Educ Sci. 2021;11(11):699. https://doi.org/10.3390/educsci11110699.
9. Bouchée T, Thurlings M, de Putter-Smits L, Pepin B. Investigating teachers' and students' experiences of quantum physics lessons: opportunities and challenges. Res Sci Technol Educ. 2023;41(2):777–99. https://doi.org/10.1080/02635143.2021.1948826.
10. Rodriguez LV, van der Veen JT, Anjewierden A, van den Berg E, de Jong T. Designing inquiry-based learning environments for quantum physics education in secondary schools. Phys Educ. 2020;55(6):065026. https://doi.org/10.1088/1361-6552/abb346.
11. Bouchée T, de Putter-Smits L, Thurlings M, Pepin B. Towards a better understanding of conceptual difficulties in introductory quantum physics courses. Stud Sci Educ. 2022;58(2):183–202. https://doi.org/10.1080/03057267.2021.1963579.
12. Chiofalo ML, Foti C, Michelini M, Santi L, Stefanel A. Games for teaching/learning quantum mechanics: a pilot study with high-school students. Educ Sci. 2022;12(7):446. https://doi.org/10.3390/educsci12070446.
13. Shi J, Fu X, Li H. Teaching suggestions of quantum information to undergraduate students. High Educ Res. 2021;6(5):142–7. https://doi.org/10.11648/j.her.20210605.18.
14. Seskir ZC, Migdał P, Weidner C, et al. Quantum games and interactive tools for quantum technologies outreach and education. Opt Eng. 2022;61(8):081809.
15. McDermott S, Vadla S, Bommanapally V, Parakh A, Subramaniam M, Ostler E. Teaching quantum cryptography using a virtual 3D educator: QuaSim. Paper presented at: National Cyber Summit; June 7, 2017; Huntsville, United States of America. https://louis.uah.edu/cyber-summit/ncs2017/ncs2017papers/10.
16. López-Incera A, Hartmann A, Dür W. Encrypt me! A game-based approach to Bell inequalities and quantum cryptography. Eur J Phys. 2020;41(6):065702. https://doi.org/10.1088/1361-6404/ab9a67.
17. QuVis. Quantum key distribution simulation. QuVis. https://www.st-andrews.ac.uk/physics/quvis/simulations_html5/sims/cryptography-bb84/Quantum_Cryptography.html. Accessed May 21, 2024.
18. Vadla S, Parakh A, Chundi P, Quasim SM. A multi-dimensional quantum cryptography game for cyber security. J Colloq Inf Syst Secur Educ. 2019;6(2):19. https://cisse.info/journal/index.php/cisse/article/view/101.

19. Utama AN, Lee J, Seidler MA. A hands-on quantum cryptography workshop for pre-university students. Am J Phys. 2020;88(12):1094–102. https://doi.org/10.1119/10.0001895.
20. Pallotta F. Bringing the second quantum revolution into high school. arXiv preprint. 2022. https://doi.org/10.48550/arXiv.2206.15264.
21. Pospiech G. Quantum cryptography as an approach for teaching quantum physics. In: Teaching-learning contemporary physics: from research to practice. 2021. p. 19–31. https://doi.org/10.1007/978-3-030-78720-2_2.
22. Thorlabs. Quantum cryptography analogy demonstration kit. Thorlabs. https://www.thorlabs.de/newgrouppage9.cfm?objectgroup_id=9869. Accessed May 11, 2024.
23. Akdemir Z, Menekse M, Hosseini M, Nandi A, Furuya K. Introducing quantum key distribution to high school students. Sci Teach. 2021;88:44–51. https://eric.ed.gov/?id=EJ1283996.
24. Bungum B, Bøe MV, Henriksen EK. Quantum talk: how small-group discussions may enhance students' understanding in quantum physics. Sci Educ. 2018;102(4):856–77.
25. Tytler R, Prain V, Hubber P, Waldrip B, editors. Constructing representations to learn in science. Berlin: Springer; 2013. https://doi.org/10.1007/978-94-6209-203-7.
26. Treagust DF, Duit R, Fischer HE, editors. Multiple representations in physics education. vol. 10. Berlin: Springer; 2017. https://doi.org/10.1007/978-3-319-58914-5.
27. Rexigel E, Kuhn J, Becker S, Malone S. The more the better? A systematic review and meta-analysis of the benefits of more than two external representations in STEM education. Educ Psychol Rev. 2024;36(4):124. https://doi.org/10.1007/s10648-024-09958-y.
28. Dooren WV, Bock DD, Verschaffel L. From addition to multiplication… and back: the development of students' additive and multiplicative reasoning skills. Cogn Instr. 2010;28(3):360–81. https://doi.org/10.1080/07370008.2010.488306.
29. Küchemann S, Ubben M, Dzsotjan D, et al. The impact of an interactive visualization and simulation tool on learning quantum physics: results of an eye-tracking study. arXiv preprint. 2023. arXiv:2302.06286.
30. Thees M, Kapp S, Strzys MP, 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. https://doi.org/10.1016/j.chb.2020.106316.
31. Altmeyer K, Kapp S, Thees M, Malone S, Kuhn J, Brünken R. The use of augmented reality to foster conceptual knowledge acquisition in STEM laboratory courses—theoretical background and empirical results. Br J Educ Technol. 2020;51(3):611–28. https://doi.org/10.1111/bjet.12900.
32. Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B. Recent advances in augmented reality. IEEE Comput Graph Appl. 2001;21(6):34–47. https://doi.org/10.1109/38.963459.
33. Cao W, Yu Z. The impact of augmented reality on student attitudes, motivation, and learning achievements—a meta-analysis (2016–2023). Humanit Soc Sci Commun. 2023;10(1):1–12. https://doi.org/10.1057/s41599-023-01852-2.
34. Chang HY, Binali T, Liang JC, et al. Ten years of augmented reality in education: a meta-analysis of (quasi-) experimental studies to investigate the impact. Comput Educ. 2022;191:104641. https://doi.org/10.1016/j.compedu.2022.104641.
35. Malone S, Garzón J, Kuhn J. Three decades of augmented reality in education: a second-order meta-analysis and research synthesis. 2023. https://doi.org/10.31219/osf.io/amw4t.
36. Sun JCY, Ye SL, Yu SJ, Chiu TK. Effects of wearable hybrid AR/VR learning material on high school students' situational interest, engagement, and learning performance: the case of a physics laboratory learning environment. J Sci Educ Technol. 2023;32(1):1–12. https://doi.org/10.1007/s10956-022-10001-4.
37. Schlummer P, Abazi A, Borkamp R, et al. Seeing the unseen—enhancing and evaluating undergraduate polarization experiments with interactive mixed-reality technology. Eur J Phys. 2023;44(6):065701. https://doi.org/10.1088/1361-6404/acf0a7.
38. Laumann D, Schlummer P, Abazi A, et al. Analyzing the effective use of augmented reality glasses in university physics laboratory courses for the example topic of optical polarization. J Sci Educ Technol. 2024;33:668–85. https://doi.org/10.1007/s10956-024-10112-0.
39. Zable A, Hollenberg L, Velloso E, Goncalves J. Investigating immersive virtual reality as an educational tool for quantum computing. In: Proceedings of the 26th ACM symposium on virtual reality software and technology. 2020. p. 1–11. https://doi.org/10.1145/3385956.3418957.
40. Dzsotjan D, Coban A, Hoyer C, Küchemann S, Durst J, Kuhn J. Quantum cryptography visualized: assessing visual attention on multiple representations with eye tracking in an AR-enhanced quantum cryptography student experiment. arXiv preprint. 2024. https://doi.org/10.48550/arXiv.2410.21975.
41. Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ. 2023;103:102274. https://doi.org/10.1016/j.lindif.2023.102274.
42. Steinert S, Avila KE, Ruzika S, Kuhn J, Küchemann S. Harnessing large language models to enhance self-regulated learning via formative feedback. arXiv preprint. 2023. https://doi.org/10.48550/arXiv.2311.13984.
43. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81–112. https://doi.org/10.3102/003465430298487.
44. Wisniewski B, Zierer K, Hattie J. The power of feedback revisited: a meta-analysis of educational feedback research. Front Psychol. 2020;10:487662. https://doi.org/10.3389/fpsyg.2019.03087.
45. Narciss S, Huth K. How to design informative tutoring feedback for multimedia learning. Instr Design Multimed Learn. 2004;181195.
46. Koenka AC, Anderman EM. Personalized feedback as a strategy for improving motivation and performance among middle school students. Middle Sch J. 2019;50(5):15–22. https://doi.org/10.1080/00940771.2019.1674768.
47. Abdurrahman A, Saregar A, Umam R. The effect of feedback as soft scaffolding on ongoing assessment toward the quantum physics concept mastery of the prospective physics teachers. J Pendidikan IPA Indonesia. 2018;7:41–7. https://doi.org/10.15294/jpii.v6i2.7239.
48. Winstone NE, Boud D. The need to disentangle assessment and feedback in higher education. Stud High Educ. 2022;47(3):656–67. https://doi.org/10.1080/03075079.2020.1779687.
49. Henderson M, Ryan T, Phillips M. The challenges of feedback in higher education. Assess Eval High Educ. 2019;44(8):1237–52. https://doi.org/10.1080/02602938.2019.1599815.

50. Anderson RJ, Anderson R, VanDeGrift T, Wolfman S, Yasuhara K. Promoting interaction in large classes with computer-mediated feedback. In: Designing for change in networked learning environments. Proceedings of the international conference on computer support for collaborative learning 2003. 2003. p. 119–23. https://doi.org/10.1007/978-94-017-0195-2_16.
51. Young KR, Schaffer HE, James JB, Gallardo-Williams MT. Tired of failing students? Improving student learning using detailed and automated individualized feedback in a large introductory science course. Innov High Educ. 2021;46(2):133–51. https://doi.org/10.1007/s10755-020-09527-5.
52. Uchiyama S, Umemura K, Morita Y. Large language model-based system to provide immediate feedback to students in flipped classroom preparation learning. arXiv preprint. 2023. arXiv:2307.11388.
53. Jansen T, Höft L, Bahr L, et al. Empirische arbeit: comparing generative AI and expert feedback to students' writing: insights from student teachers. Psychol Erzieh Unterr. 2024;71(2):80–92.
54. Steiss J, Tate TP, Graham S, et al. Comparing the quality of human and ChatGPT feedback on students' writing. OSF preprints 2023. https://doi.org/10.1016/j.learninstruc.2024.101894.
55. Jacobsen LJ, Weber KE. The promises and pitfalls of ChatGPT as a feedback provider in higher education: an exploratory study of prompt engineering and the quality of AI-driven feedback. 2023. https://doi.org/10.31219/osf.io/cr257.
56. Dai W, Lin J, Jin H, et al. Can large language models provide feedback to students? A case study on ChatGPT. In: 2023 IEEE international conference on advanced learning technologies (ICALT). 2023. p. 323–5. https://doi.org/10.1109/ICALT58122.2023.00100.
57. Agricola BT, Prins FJ, Sluijsmans DM. Impact of feedback request forms and verbal feedback on higher education students' feedback perception, self-efficacy, and motivation. Assess Educ. 2020;27(1):6–25. https://doi.org/10.1080/0969594X.2019.1688764.
58. Merry S, Orsmond P. Students' attitudes to and usage of academic feedback provided via audio files. Biosci Educ. 2008;11(1):1–11. https://doi.org/10.3108/beej.11.3.
59. Van der Schaaf M, Baartman L, Prins F, Oosterbaan A, Schaap H. Feedback dialogues that stimulate students' reflective thinking. Scand J Educ Res. 2013;57(3):227–45. https://doi.org/10.1080/00313831.2011.628693.
60. Just MA, Carpenter PA. A theory of reading: from eye fixations to comprehension. Psychol Rev. 1980;87(4):329. https://doi.org/10.1037/0033-295X.87.4.329.
61. Stark L, Korbach A, Brünken R, Park B. Measuring (meta) cognitive processes in multimedia learning: matching eye tracking metrics and think-aloud protocols in case of seductive details. J Comput Assist Learn. 2024;40(6):2985–3004.
62. Lai ML, Tsai MJ, Yang FY, et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ Res Rev. 2013;10:90–115. https://doi.org/10.1016/j.edurev.2013.10.001.
63. Alemdag E, Cagiltay K. A systematic review of eye tracking research on multimedia learning. Comput Educ. 2018;125:413–28. www.learntechlib.org/p/200849/.
64. Deng R, Gao Y. A review of eye tracking research on video-based learning. Educ Inf Technol. 2023;28(6):7671–702. https://doi.org/10.1007/s10639-022-11486-7.
65. Ke F, Liu R, Sokolikj Z, Dahlstrom-Hakki I, Israel M. Using eye-tracking in education: review of empirical research and technology. Educ Technol Res Dev. 2024;72:1383–418. https://doi.org/10.1007/s11423-024-10342-4.
66. Hahn L, Klein P. Eye tracking in physics education research: a systematic literature review. Phys Rev Phys Educ Res. 2022;18(1):013102.
67. Hoyer C, Girwidz R. Animation and interactivity in computer-based physics experiments to support the documentation of measured vector quantities in diagrams: an eye tracking study. Phys Rev Phys Educ Res. 2020;16(2):020124. https://doi.org/10.1103/PhysRevPhysEducRes.16.020124.
68. Klein P, Viiri J, Mozaffari S, Dengel A, Kuhn J. Instruction-based clinical eye-tracking study on the visual interpretation of divergence: how do students look at vector field plots? Phys Rev Phys Educ Res. 2018;14(1):010116. https://doi.org/10.1103/PhysRevPhysEducRes.14.010116.
69. Klein P, Viiri J, Kuhn J. Visual cues improve students' understanding of divergence and curl: evidence from eye movements during reading and problem solving. Phys Rev Phys Educ Res. 2019;15(1):010126. https://doi.org/10.1103/PhysRevPhysEducRes.15.010126.
70. Klein P, Küchemann S, Brückner S, Zlatkin-Troitschanskaia O, Kuhn J. Student understanding of graph slope and area under a curve: a replication study comparing first-year physics and economics students. Phys Rev Phys Educ Res. 2019;15(2):020116. https://doi.org/10.1103/PhysRevPhysEducRes.15.020116.
71. Brückner S, Zlatkin-Troitschanskaia O, Küchemann S, Klein P, Kuhn J. Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: a postreplication study. Front Psychol. 2020;11:2090.
72. Küchemann S, Klein P, Becker S, Kumari N, Kuhn J. Classification of students' conceptual understanding in STEM education using their visual attention distributions: a comparison of three machine-learning approaches. CSEDU. 2020;1:36–46.
73. Ruf V, Horrer A, Berndt M, et al. A literature review comparing experts' and non-experts' visual processing of graphs during problem-solving and learning. Educ Sci. 2023;13(2):216.
74. Holmqvist K, Andersson R. Eye tracking: a comprehensive guide to methods. Paradigms and measures. 2017.
75. Kapp S, Barz M, Mukhametov S, Sonntag D, Kuhn J. ARETT: augmented reality eye tracking toolkit for head mounted displays. Sensors. 2021;21(6):2234. https://doi.org/10.3390/s21062234.
76. Dzsotjan D, Ludwig-Petsch K, Mukhametov S, Ishimaru S, Küchemann S, Kuhn J. The predictive power of eye-tracking data in an interactive AR learning environment. In: Adjunct proceedings of the 2021 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2021 ACM international symposium on wearable computers. 2021. p. 467–71. https://doi.org/10.1145/3460418.3479358.
77. Brunner K, Obersteiner A, Leuders T. How prospective teachers detect potential difficulties in mathematical tasks–an eye tracking study. Res Subj-matter Teach Learn. 2021;4(1):109–26.
78. Knoblich G, Ohlsson S, Raney GE. An eye movement study of insight problem solving. Mem Cogn. 2001;29(7):1000–9.
79. Krupinski EA, Tillack AA, Richter L, et al. Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. Hum Pathol. 2006;37(12):1543–56.

80. Holmqvist K, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J. Eye tracking: a comprehensive guide to methods and measures. London: Oxford University Press; 2011.
81. Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report. arXiv preprint. 2024. arXiv:2303.08774.
82. Shi D, Ye T. Behavioral carry-over effect and power consideration in crossover trials. Biometrics. 2024;80(2):ujae023. https://doi.org/10.1093/biomtc/ujae023.
83. Connelly LM. Understanding crossover design. Medsurg Nursing. 2014;23(4):267.
84. Cronbach LJ, Snow RE. Aptitudes and instructional methods: a handbook for research on interactions. New York: Ardent Media; 1981.
85. De Jong T, Linn MC, Zacharia ZC. Physical and virtual laboratories in science and engineering education. Science. 2013;340(6130):305–8. https://doi.org/10.1126/science.1230579.
86. Flegr S, Kuhn J, Scheiter K. When the whole is greater than the sum of its parts: combining real and virtual experiments in science education. Comput Educ. 2023;197:104745. https://doi.org/10.1016/j.compedu.2023.104745.
87. Kapici HO, Akcay H, de Jong T. Using hands-on and virtual laboratories alone or together–which works better for acquiring knowledge and skills? J Sci Educ Technol. 2019;28:231–50.
88. Dahlan JA, Wibisono Y. The effect of hands-on and computer-based learning activities on conceptual understanding and mathematical reasoning. Int J Instr. 2021;14(1):143–60.
89. Schwichow M, Zimmerman C, Croker S, Härtig H. What students learn from hands-on activities. J Res Sci Teach. 2016;53(7):980–1002. https://doi.org/10.1002/tea.21320.
90. Lee H, Hsia CC, Tsoy A, Choi S, Hou H, Ni S. VisionARy: exploratory research on contextual language learning using AR glasses with ChatGPT. In: Proceedings of the 15th biannual conference of the Italian SIGCHI chapter. 2023. p. 1–6. https://doi.org/10.1145/3605390.3605400.
91. Lau KHC, Bozkir E, Gao H, Kasneci E. Evaluating usability and engagement of large language models in virtual reality for traditional Scottish curling. arXiv preprint. 2024. https://doi.org/10.48550/arXiv.2408.09285.
92. Gao H, Huai H, Yildiz-Degirmenci S, Bannert M, Kasneci E. DataliVR: transformation of data literacy education through virtual reality with ChatGPT-powered enhancements. arXiv preprint. 2024. https://doi.org/10.48550/arXiv.2409.08738.
93. Chheang V, Sharmin S, Márquez-Hernández R, et al. Towards anatomy education with generative AI-based virtual assistants in immersive virtual reality environments. In: 2024 IEEE international conference on artificial intelligence and eXtended and virtual reality (AIxVR). 2024. p. 21–30. https://doi.org/10.1109/AIxVR59861.2024.00011.
94. Nuswantari A, Wu YT, Surjono HD. System usability scale measurement on synchronous online argumentation learning system. In: International conference on online and blended learning 2019 (ICOBL 2019). 2020. p. 142–6. https://doi.org/10.2991/assehr.k.200521.031.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com