How can an educational chatbot's feedback influence human attention?

Francesca Sylvester [1], Nguyen-Thinh Le [1,2] * and Surjo R. Soekadar [2]

*Correspondence:
nguyen-thinh.le@charite.de
Humboldt-Universität zu Berlin,
Germany
Charité – Universitätsmedizin
Berlin, Germany

Abstract

Educational chatbots have been shown to be useful assistants in computer-supported learning settings. However, how does feedback of an educational chatbot impact on the learner's attention? Thus, this paper proposes a study to measure changes in human attention when learning by using an educational chatbot Liza that is intended to improve human reasoning ability. In total, 18 participants participated in the study and had a conversation with Liza. During the interaction with Liza, the attention of the study participants was measured using a mobile electroencephalogram (EEG) device. Three findings have been determined based on statistics methods. First, it was found that there was a significant attention effect occurring in 54% of the times, after the educational chatbot showed feedback and the attention measurement took place over the length of a task. Second, when differentiating the type of feedback, positive feedback had a significant effect in 71 of these 137 cases (51.82%) and negative feedback had a significant effect in 66 of the 137 cases (48.18%). Third, statistical results showed that there was no significant difference in attention at the significant level of 0.05 during the 10 seconds before, and 10 seconds after positive feedback is received. Similar is the case for negative feedback.

Keywords: Attention, Pedagogical agent, Feedback, EEG, Physiological computing

Introduction

In 2020, the COVID-19 pandemic disrupted the world. A majority of communication suddenly had to take place digitally, whether it was over e-mail, phone calls or video calls, reducing in-person contact to stop the spreading of the virus. This also meant that secondary and tertiary education had to switch from a traditional face-to-face setting to distance learning. Teachers had to adjust to the new challenges that came with online learning. On the other hand, students also struggled with the new form of learning settings, facing more distractions when they are in the comfort of their own home, as well as the

overall deterioration of performance that comes with digital learning environments (Junco & Cotton, 2012). Teachers therefore need to ensure that the students pay full attention to the online lesson, since attention has been identified as one of the key factors for learning performance, aiding the learning process by helping students to focus on important information (Al'Omari & Balushi, 2015, p. 685).

How can teachers in an online learning setting determine the learner's attention? In order to determine whether a learner is attentive, there have been successful attempts that measured the physiological parameters of students with an electroencephalogram (EEG), deriving the level of attention (Chiang et al., 2018). While it certainly is possible to implement such an attention measurement device in an educational chatbot, the measurement of attention takes multiple steps and can be rather costly, making it an inadequate choice for an educational environment. Therefore, to "wake up" the attention and thus, the learning performance of a student, another useful instrument could be giving adaptive feedback during learning.

Teachers give feedback during a lesson, knowing of the importance of feedback in education since it can double the average student performance (Hattie & Timperley, 2007, p. 83). From the technical point of view, an educational chatbot may also adapts the pedagogical strategy of teachers and should give feedback at appropriate time points. A research gap is whether an educational chatbot's feedback would possibly influence the attention of a learner and thus, her learning performance. This paper proposes a research question: How can an educational chatbot's feedback influence human attention?

For this study, an experiment is conducted, in which participants talk to an educational chatbot that has been developed to improve human reasoning. During the entire conversation, the participant solves numerous tasks while the attention is being monitored with a mobile EEG device. After each solution, the educational chatbot responds with feedback, letting the participant know if they solved the answer correctly or falsely.

State of the art

The concept of attention

The definition of attention may have been first coined by William James in 1890 as follows:

"*It is the taking possession by the mind, in clear and vivid form, of one out of what seems several simultaneously possible objects or trains […] of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed scatterbrained state […].*" (James, 1890).

Later on, the definition of attention has been differentiated in four main types (McDowd et al., 1991): Sustained attention, selective attention, alternating attention and divided

attention. Sustained attention, also known as "vigilance" or "tonic alertness" pertains to maintaining focus with a moderate level of mental effort over an extended period of time (Oken et al., 2006; McDowd, 2007).

Selective attention is the process of actively selecting focus on one stimulus, including the external environment or internal sources, while filtering out others (Johnston & Dark, 1986). A famous example of selective attention is the cocktail party effect by Colin Cherry in 1953, which describes the ability to focus on one conversation while multiple are happening around (Lindsay, 2020, p. 2).

Alternating attention is the ability to switch back and forth between tasks that require different cognitive processes (Sohlberg & Mateer, 1987).

Finally, divided attention, commonly known as "multi-tasking", is the activity of processing more than one stimulus at one time or reacting to multiple stimuli simultaneously. According to the definition for different types of attention, sustained attention is most highly related to learning because learners need to maintain their focus on a specific learning context over a period of time (Esterman & Rothlein, 2019, p. 174).

Attention in online learning

Maintaining attention is crucial to learning. Various studies demonstrated that attention has an impact on achieving a better learning performance and thus being essential in a learning environment (Al'Omari & Balushi, 2015; O'Connell et al., 2009). However, whereas computer-supported learning technologies come with a considerable number of benefits, it is undeniable that they also come with disadvantages. The downside is the sheer number of possible distractions, e.g., a notification of a new e-mail appearing on the screen, while a student is using a computer. Through an online survey, Junco and Cotton (2012) showed that the more college students engage with digital learning environments, the worse their performance will be. In a meta-analysis, Delgado et al. (2018) reviewed a research that compared reading on paper to reading on digital screens: the census was a clear advantage for paper, discussing that digital environments are not suitable for deep comprehension and learning. Since most computer-supported learning environments rely on quick interactions, students might find it hard to maintain their attention for a reading comprehension on a digital device (Delgado et al., 2018, p. 20). Also Lodge and Harrison (2019) analyze the role of attention in the online learning environment and claimed that the easy accessibility of information negatively impacts the capacity to learn.

Although the lack of attention has been identified as a side-effect of online learning, and thus, impacts negatively on the learning performance, both students and teachers perceive online learning as useful environments (Liaw et al., 2007).

Attention and feedback

Feedback is a kind of information we receive after an action has been performed (Nelson & Schunn, 2008). In a learning setting, feedback can be given by a teacher or a classmate.

Feedback can be divided into two categories: positive and negative feedback. Positive feedback is received when an action has been performed correctly, vice versa, negative feedback is given. Feedback resulting in learning could be explained by the behaviorism theory. One of the pioneers of the behaviorism learning theory was Thorndike (1927) who asked subjects in an experiment to draw a four-inch line while being blindfolded. He discovered that when not receiving any sort of feedback, the subjects did not improve and therefore have not learned; whereas subjects that have received feedback (simply stating if their drawing is wrong or right) gained a considerable accuracy for drawing lines.

Furthermore, research focuses on the impact of positive or negative feedback on the learning performance. Brackbill and O'Hara (1958) successfully test their hypothesis, that humans learn faster when receiving both reward (positive feedback) and punishment (negative feedback), rather than just reward. However, this study did not include a punishment-only control group. Meyer and Offenbach (1962) found that the punishment-only (with negative) control group led to a faster improvement in learning performance.

While some studies (Brackbill & O'Hara, 1958; Meyer & Offenbach, 1962) show the effect of feedback on learning to be overall improving, with negative feedback having a greater impact, Kluger and DeNisi (1996) criticize this assumption. In a meta-analysis study, Kluger and DeNisi (1996) find that feedback has a high variability of effects, depending on types of learning and continuity of feedback intervention.

Human attention using educational technologies

Research studies with regard to human attention using educational technologies pursue mainly two goals (Table 1): 1) supporting the teacher with an attention recognition system, and 2) observing changes of attention during a lecture.

Arbel and colleagues (2020) conduct a study with twenty young adults who had to perform a variation of learning tasks while their attention was measured via eye-tracking. Any change of attention was calculated through the two main eye-gaze's measures: proportion fixation time and fixation probability. If a task was solved correctly, the subject received visual positive feedback of three green checks; if solved incorrectly, three red Xs appeared. In their study, they concluded that attention of the subjects changed when feedback occurred, with the change being greater when negative feedback has been received (Arbel et al., 2020, p. 6).

Liu and colleagues (2013) build a system supporting the teacher in recognizing the level of student's attention in distance learning, using data from a mobile EEG device. Zalatelj and Košir (2017) pursue a similar research, however, in traditional classroom settings,

Table 1 Student's attention in educational environments

Aim	Procedure	Participants	Results	Ref.
Recognizing attention levels using mobile EEG sensors	Subjects had to listen to English phrases and answer related questions in two phases, while attention was measure. Support Vector Machine (SVM) was used to identify an attentive state. 4 Phases (Resting Phase, Interaction Phase without Interference, Interaction Phase with Interference, Questionnaire)	24 students (12 male, 12 female)	SVM model recognizes the attention level with classification accuracy of 76.82%	Liu et al. (2013)
Using a motion sensor to predict attention in a face-to-face setting with a machine learning algorithm	Sensor recorded the students during an entire lesson; after-wards, recorded data was analyzed by humans to derive a set of body features indication attentive state; then seven machine learning algorithms were trained to calculate attention score	22 students (20 male, 2 female)	Created attention monitoring system able to predict student's attention with an accuracy of up to 75.3%	Zaletelj & Košir (2017)
Using mobile EEG device to observe the possibility of multiple readings of attention level in a classroom	The students were instructed by the researchers about the device and video cameras were put up to record the lesson for later interpretation. After the preparations, the relevant data was recorded while the students participated in a class. 2 Phases (Preparation Phase, Recording Phase)	42 students	Certain teaching material increased attention level, such as PowerPoint presentations and digital maps; device found to be usable for multiple measures	Sezer et al. (2015)
Measurement of the attention level with a mobile EEG device when reading on different kinds of media	Students were given a tablet that contained three different media types (text, text and graphic, video) for learning with a similar topic; students then had to use each media type and answer a questionnaire about the topic after each reading. 7 Phases (Introduction Phase, Resting Phase, 1st Reading, Resting Phase, 2nd Reading, Resting Phase, 3rd Reading)	28 students (19 male, 7 female)	Greater attention in text media, but no statistical significance; students that were experienced with tablets showed stable attention	Ni et al. (2020)
Observing the attention decline of students during a lecture with response devices (self-report)	Students had to use a response device (clicker), reporting lapses of attention throughout a lecture by pressing one of three buttons that represented the length they did not pay attention (1 minute, 2-3 minutes, 5 or more minutes)	186 students	No continuous attention for entire lecture but rather attention fluctuates from engaged to non-engaged in short cycles	Bunce et al. (2010)

estimating a student's attention level by monitoring their body posture and gaze during a lecture with the help of a motion sensor. Even though different measurement methods of attention have been used, both recognition systems (Liu et al., 2013; Zalatelj & Košir, 2017) have a similar accuracy with regard to identifying attentive states.

Other researchers attempt to compare students' attention in computer-supported learning against text-based learning. Whereas two research groups find attention is higher in teaching methods that either include a tool such as PowerPoint or digital maps (Sezer et al., 2015), or a chemical demonstration (Bunce et al., 2010). Ni et al. (2020) suggest that attention is greater when students learn with text-only material compared to material that contained videos or graphs.

Only one relevant study that measures attention values while a participant is talking to a chatbot was found. Bitner and Le (2021) validate the algorithm of a mobile EEG device to determine attention of a subject through the method of an experiment. In the study, 27 participants were instructed to interact with the pedagogical agent Synja, teaching the user concepts of the programming language Java, while simultaneously wearing the mobile EEG device. The authors report only weak evidence with only 5 participants showing a higher attention value when answering correctly. In addition, the study suggests that the average attention score leading up to a correctly answered question is higher than when answering a question incorrectly (Bitner & Le 2021).

Methods to measure attention

In a traditional classroom, attention is often differentiated by on-topic and off-topic activities (Keller et al., 2020). Teachers may recognize if a student's attention is on-topic or off-topic by observing the body language of their students. While there are some clear signals that point to on-topic attention, such as a student nodding in response to a teacher's presentation, other signals may indicate to both on-topic and off-topic behaviors. For example, a student staring outside the window can either mean, that their attention is off-topic, or that they're thinking about a question the teacher just asked (Keller et al., 2020). It is thus up to teacher, to interpret these signals correctly. When an educational chatbot takes the role of a teacher, different measurement methods can be used to predict the attention's level of a student.

There are qualitative and quantitative measurement methods of attention. The focus of this paper will be the quantitative methods that can be categorized into indirect and direct types. The indirect measurement relies on observing at an individual, e.g., body movement and eye-tracking. Eye-tracking is the "most widely used tool for measuring visual attention" (Mancas & Ferrera, 2016, p. 22). The direct measurement method of attention is based on physiological responses of individuals, e.g., brain signals. The human brain runs our body functions and our thinking. When the brain tells our body what to do, millions of neurons

(brain cells) are activated, producing a current flow (Teplan, 2002). In order to measure electrical brain activity, electroencephalogram (EEG) can be applied. Multiple electrodes are placed on the scalp of a subject in order to record electric flows through the brain. There are five widely recognized brain waves, which are differentiated by their unique frequency range (Teplan, 2002): Delta δ (0.5-4 Hz), Theta θ (4-8 Hz), Alpha α (8-13 Hz), Beta β (13-30 Hz), and Gamma γ (>30 Hz).

Studies have proven the correlation between sustained attention with a decrease of Alpha wave activity. Ray and Cole (1985) conduct an experiment involving college students, who are asked to complete several tasks over a course of three days. The study reports higher parietal Alpha in tasks that do not require attention compared to those that do, thus a decrease of Alpha during attentional tasks. O'Connell and colleagues (2009) also conclude that an increase in Alpha band activity corresponds to a decrease of attention. In a rather recent study, the widely accepted correlation of Alpha waves and attention has been investigated again and the authors rightfully claim that "correlation is not causality" (Bagherzadeh et al., 2020, p. 577). In the study, the subjects were divided into two groups, one that learned how to manipulate the Alpha power in the left hemisphere and the other one the right hemisphere. After completing a task, the researchers found that a higher Alpha power in the left-group led to a higher attention over the right-group, thus supporting the causal relationship of attention and Alpha synchrony.

While Alpha band power is agreed to be correlated with attention, other EEG band powers are controversially found to be in relation with attention. Aftanas and Golocheikines (2001) suggests a link between attention and Theta activity. By observing the EEG during different mediation phases, they found that subjects with little experience in meditation had an absence of Theta activity, since they may have anxious feelings towards not maintaining a meditative state, thus having a narrow attentional focus.

Pavlygina et al. (2011) also report that Delta activity increases while problem-solving and thus, an increase of attention. Contrary to the previous finding regarding Theta activity, Oken et al. (2006) find that there are two measures that often correlate with a bad task performance: increasing Theta and decreasing Beta. This result that decrease in Beta activity during inattentive behavior is in accordance with the report of Linden et al. (1996).

Methodology

Since there is a research gap on the effect of feedback given by a learning system regarding learner's attention, the research question "How can an educational chatbot's feedback influence human attention?" is of high relevance. Regarding the possible effect of feedback on attention, the following hypotheses will be tested:

H_0: Neither feedback nor negative feedback has an impact on human attention.

H_1: Positive feedback will lower the level of attention while negative feedback will improve it.

The H_1 hypothesis is supported by the research of Arbel et al. (2020) in psychology. To reject either the H_0 or H_1 hypotheses, it is mandatory to measure two variables: the feedback of the educational chatbot and the human attention of a person that is talking to the chatbot. To measure these two variables, an experiment must be carried out.

Materials

The experiment consists of a conversation between the educational chatbot LIZA and study participants in a laboratory environment, while the participant wears NeuroSky's MindWave Mobile 2. The attention levels (10, 20, 30, 40, 50, 60, 70, 80, 90, 100) of the user are provided using the Attention Meter prebuilt algorithm of NeuroSky[1].
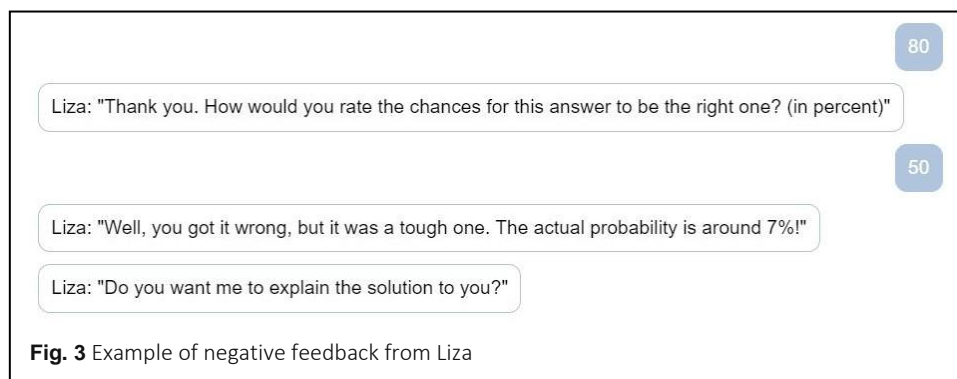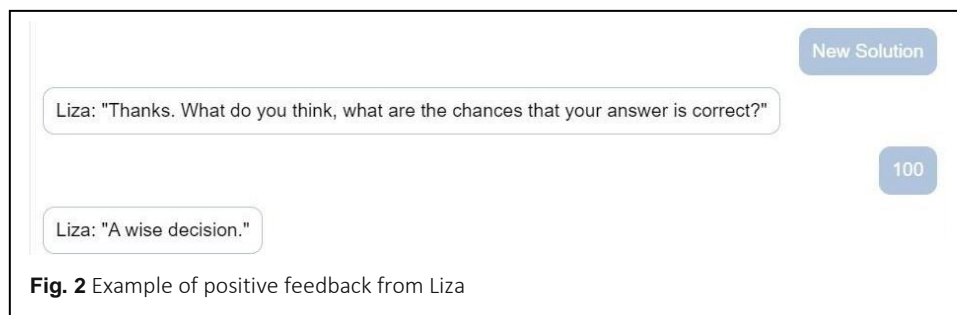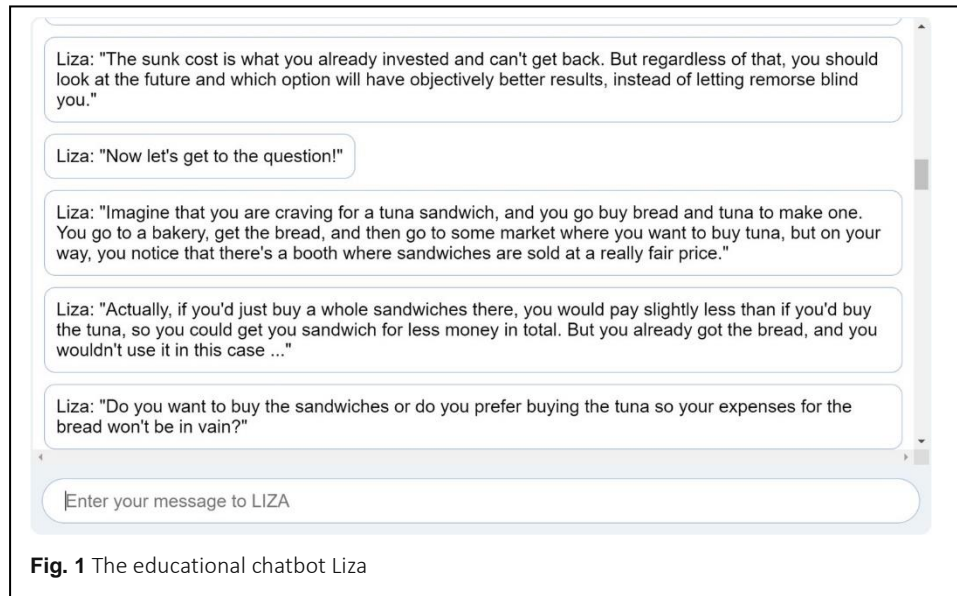
Educational chatbot: LIZA

Wartschinski et al. (2017, p. 249) developed Liza because teaching human reasoning skills requires repeated explanations and examples, and since the resources of a human teacher are limited and this topic is not usually offered at school, an educational chatbot can serve this purpose for teachers.

The web-based educational chatbot Liza[2] was programmed using Java. The knowledge base is referred to as "Story Store" and contains the human reasoning tasks that are categorized in seven topics. The seven topics are: Bayesian reasoning, the Law of Large Numbers, the Gambler's Fallacy, Wason's Selection Task, Covariance Detection, the Sunk Cost Fallacy, and Belief Bias in Syllogistic Reasoning. The tasks were taken from well-known psychology studies that have shown validity in literature (Wartschinski et al., 2017). A lesson begins with Liza greeting the user and introducing herself and her aim. Afterwards she starts the training phase by first announcing one of the seven topics with a short explanation (if the user wants to have an explanation), followed by an example case the user has to solve (see Figure 1).

If the task has been solved correctly, Liza gives the user a short, positive, and enthusiastic response (Figure 2). However, if solved falsely, Liza responds in a concerning way, nonetheless still encouraging the user (Figure 3).

At the end of the conversation, after 14 human reasoning tasks have been solved (two for each reasoning topic), Liza gives the user a summary of their conversation, giving a final feedback on the user's abilities. Liza has been thoroughly designed and programmed to overcome the mentioned common design challenges, mainly working task-oriented while using common social behavior cues (Wartschinski et al., 2017, p. 251). Upon the user's request, Liza gives hints to solve the task. If a task has been solved incorrectly, she offers

Fig. 1 The educational chatbot Liza

Fig. 2 Example of positive feedback from Liza

Fig. 3 Example of negative feedback from Liza

explanation to improve the solution. To seem less like a computer, she takes the role of a trustworthy mentor and tries to have small off-topic conversations with the user, for example, asking if the user has ever been to a different city. However, most of the off-topic conversation remains close to the actual content of the lesson, as Liza uses the user's response for an upcoming task. If the user replies "Yes, I have been to Paris!", Liza remembers that answer and might propose a task, that takes place in Paris. Liza aims to show emotions like pride, joy, and curiosity, as well as disappointment, sadness, and insecurity (Wartschinski et al., 2017, p. 252). An evaluation study for Liza was conducted, in which the participants were divided into a treatment and a control group; the treatment group was talking to the agent, whereas the control group only read short texts about each topic. After conducting the experiment, the results showed that the treatment group had an overall better performance than the control group; the participants also noted that they liked working with Liza and had a positive experience with her. For her success in teaching, high usability and implemented feedback algorithm, Liza was chosen for the role as the educational chatbot in this paper.

For the study being presented, a different version of Liza has been developed. This version was programmed in Python with the help of RASA[3], an open-source machine learning framework for conversational chatbots. Instead of seven topics, the simpler version of Liza for this study only provides four topics (Sunk Cost Fallacy, Gamblers Fallacy, Regression to the Mean and Base Rate Fallacy). The following examples and explanations are directly taken from Liza's knowledge base and serve as an understanding of what exactly Liza teaches, and what the participants had to solve. All reasoning tasks have been validated in psychology studies and based on rational thinking composite by Toplak et al. (2014) and the collection of rational tasks of Larrick (2004).

The Sunk Cost Fallacy describes the tendency to continue investing time or money, not recognizing a bad outcome. For example, Liza might create a situation where the user accidentally bought tickets to two different plays for the same day; not being able to return the tickets, the user must decide which play they want to see: the one that costs more but is less entertaining; or the cheaper one, that sounds more promising. The right decision is the second one since the money has been already spent and cannot be retrieved. However, since it is a fallacy, most people would choose to go to the more expensive play.

The Gambler's Fallacy occurs when people observe a series of events and think that the events that occurred more frequently will appear less frequent in the future. To improve the users' skills, Liza will explain a situation of past events, and asks the user what the next event might be. To make this less abstract: Liza explains a situation, in which a coin was tossed a few times. Each time, the coin showed the same side. For the next toss, Liza asks the user which side (head or tails) will most likely appear. According to the Gambler's

Fallacy, the user will say tails since heads has been leading in the previous events. The right answer however is 50:50, because past events cannot change future events.

Regression to the Mean is the effect an extraordinary sample has on the understanding of a whole population. It happens when a large or small sample occurs, followed by a sample that is closer to the average of the population. For example, there are two hospitals, one having more births per day than the other (100 vs. 10). Liza then asks the user, which hospital is more likely to have higher rate of male births on a specific day. Most people would assume the answer to be the hospital that has 100 births per day, even though the right answer is the hospital with less births, since having more births means that a great deviation is less likely to happen.

The last reasoning topic is the Base Rate Fallacy and occurs during problems with conditional probabilities. To train this complex reasoning topic, the users are calculating the chances of a certain event happening. For example, the user must solve a problem where they determine the chance of a person having a specific disease upon receiving a positive test result. Liza gives the user numbers about the occurrence of the disease in the population and the false-positive/false-negative-rate to help the user solve the problem. Most people tend to find these sorts of tasks rather difficult and therefore solve it incorrectly. The reason is that people ignore parts of statistical information as it confuses them.

In the version implemented for this study, for each of four reasoning topics, four reasoning tasks were generated. Therefore, Liza will give 16 tasks (four for each reasoning topic) in total. The four tasks of each reasoning topic are not asked successively, but alternatively: a Sunk-Cost-Fallacy task is followed by a task of Gamblers Fallacy, which is followed by a task of Regression to the Mean, which is lastly followed by a Base Rate Fallacy task. After this, the cycle is repeated three more times.

Also the language used by Liza has been translated into German in the new version (see Figure 4), because the target group of study participants would be native German speakers and the focus of attention will be on the learning context, not the language. This is also supported by a study conducted by Liu and colleagues (2013), in which the attention of students during a lecture was recorded. The students were reading a text in English, which was not their native language. The study found that the Delta waves were more active during the times, when the reading took place (Liu et al., 2013). Since an increase of Delta activity positively correlates to increasing attention, any factors that could lead to a misinterpretation of results were removed.

The research question being investigated can only be properly evaluated if the subject receives both positive and negative feedback. Since the four reasoning topics vary in their difficulty, it can therefore be assumed, that every participant will receive both positive and negative feedback, which makes Liza an ideal educational chatbot in this study.

Fig. 4 Liza's user interface
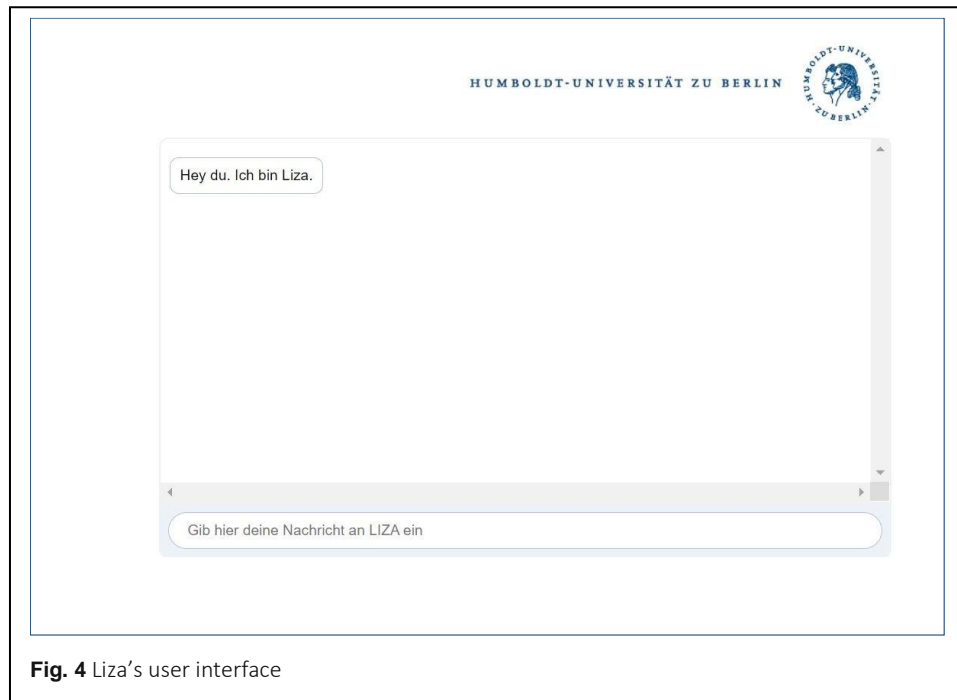
NeuroSky's MindWave Mobile 2

Due to its simple usability and accuracy, the MindWave[4] Mobile 2 headset had been used to record the attention of the study participants. The headset's reference and ground electrodes are on the ear clip and the EEG electrode is on the sensor arm. The EEG electrode is put on the forehead above the eye, i.e., the FP1 position of the frontal lobe (Teplan, 2002).

The MindWave Mobile 2 uses Bluetooth to connect to other devices, in this case a laptop. To collect the actual physiological data, OpenVibe[5] has been used. OpenVibe is an open-source software for Brain-Computer-Interfaces and can be used to acquire, filter, and analyse brain signals at real-time. The signal acquisition consists of multiple steps. After connecting the MindWave Mobile 2 with the laptop via Bluetooth, it needs to be selected by the OpenVibe Acquisition Server. Here one can choose the age and gender of the subject, as well as what kind of physiological parameter should be measured, e.g., attention. The whole EEG power spectrum (low and high Alpha; low and high Beta; Gamma and mid Gamma; Delta and Theta) are recorded. The MindWave Mobile 2 collects data at a frequency of 512 Hz, however both the eSense Attention Meter algorithm and the power spectrum are only calculated at a frequency of 1 Hz. Once this is done and the recording is started, the OpenVibe Designer (Figure 5) comes into play. The OpenVibe Designer is where the processing pipeline (e.g., filtering or visualisation) of the brain signals takes place. For this study, the following components were used: an acquisition client, that

Fig. 5 OpenVibe
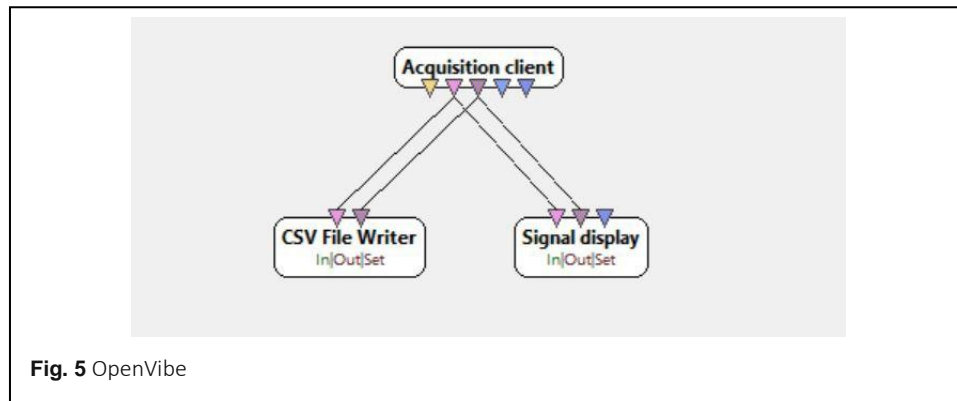
retrieves data from the acquisition server, a signal display that shows the recorded data, and a csv-file writer, in which all the recorded signals were exported to after the conversation with Liza ended.

Participants

Human reasoning affects everyone and is not limited to a specific group of students or persons in general. Therefore, almost anyone can participate in this experiment. When it comes to attention and age, there is an assumed and proven decline of attention with a higher age (Zanto et al., 2011). It was therefore necessary to set an age limit. While various research find a decline in sustained attention when comparing very young people to old people, e.g., Zanto et al. (2011), a recent study took a deeper look at more specific age groups (Lufi & Haimov, 2018). Lufi and Haimov (2018) choose 496 participants, divided into eight age groups, spanning ten years; the youngest group started at 12 years. The participants had to complete a computerized math test, that assessed attention. In their results, they found that there was an increase of attention in the first three age groups (12 – 39.99), and a steady decline after that. The peak attention level has been found in the age group 30 – 40 (Lufi & Haimov, 2018). Therefore, the age limit for this experiment was set at 40. In order to covering such an age limit of 40 years, we invited friends and family relatives to participate in the study who desired to improve reasoning skills in addition to university students.

Another important aspect when it comes to attention is possible diseases that might affect attention. The most known is attention deficit hyperactivity disorder (ADHD). ADHD comes with a "pattern of inattention […] that interferes with functioning or development" (Battle, 2013, p. 59). To identify whether a person suffers of ADHD, the Diagnostic and Statistical Manual of Mental Disorders (Battle, 2013) describes a diagnostic of nine symptoms; an adult that meets at least five of these is very likely to have ADHD. Thus, all participants had to fill out a questionnaire to identify a possible attention disorder. Having

an attention disorder would not disqualify a participant, but rather explain potential abnormalities in attention.

In total, 18 people participated in this study. There were 11 women and 7 men, ranging between 23 and 40 years of age, with an average age of 30.72. All participants finished high school and are therefore qualified to study at a university, four have a bachelor's degree and five have a masters or higher. Based on answers of the questionnaire Diagnostic and Statistical Manual of Mental Disorders, there was no indications of an attention disorder found in any of the participants.

Experiment procedure

The experiment took place at an office room in 12043 Berlin (Figure 6). Over the course of two weeks, the participants were asked to come to the building to undergo the experiment. Before each experiment, the MindWave Mobile 2 device was wiped clean with a damp cloth and a new battery was placed in to ensure that it would last for the entire session. The educational chatbot was run and ready to start a new conversation. Additionally, the room was aired for 10 minutes and anything the participants might touch was disinfected, to reduce the spread of COVID-19. The experiment itself consisted of four phases.

Introduction Phase. In this phase, the participants were introduced to the experiment and its goal, however, they were not explained about the types of attention. The participants had to read and sign a consent form agreeing to their data being used, as well as answer a pre-questionnaire. The pre-questionnaire consisted of two parts. In the first part, the

Fig. 6 Experimental setup

participants had to state their age, degree of education (completion of secondary education, bachelor's degree, master's degree or higher, or other) as well as their current employment. Following this, the participants had to answer questions to identify a possible attention disorder. This part consisted of six questions in accordance with the Diagnostic and Statistical Manual of Mental Disorders (Battle, 2013). With the help of a Likert-Scale (never, rarely, sometimes, often, very often) participants had to rate how often they have difficulties with concentrating on a task or conversation. After completing the questionnaire, the participants were further instructed on the conversation with Liza, explaining what exactly Liza teaches and to respond by simply typing the answer in the intended text field. Furthermore, since some of Liza's tasks involved mathematical problems, the participants were given a calculator, as well as a pen and paper to write down any thoughts they had while solving a task. In total, this phase lasted about 10 minutes.

Resting Phase. After introducing the participants to the experiment, the MindWave Mobile 2 device was placed on their head. Then, the OpenVibe program for recording EEG data was started. Once it was ensured that everything was running smoothly, the participants were allowed to rest for 5 minutes. We decided to have 5 minutes for resting because the short break lasts commonly between 3 to 10 minutes for sustained attention tasks. This was mainly done so the participant can get used to the new environment and to relax (Schumann et al., 2022), so that the brain activity was calmed before having a conversation with Liza. Overall, this phase lasted about 10 minutes.

Experimental Phase. In this phase, the participants had their conversation with Liza to solve reasoning tasks. Based on previous experiments with Liza (Le & Wartschinski, 2018), we estimated 60 minutes at maximum for solving the 16 reasoning tasks. The real time however depended on how quickly or slowly the participants solved the tasks.

Post-Experiment Phase. After having a conversation with the educational chatbot Liza, the participant was allowed to take the MindWave Mobile 2 device off. The participant then had to fill a post-questionnaire that consists of similar questions as in the pre-questionnaire, and questions estimating their own attention level (very low, low, average, high, very high).

Data collection and analysis methods

During the experiment, two files are generated to log relevant data: A text file contains the conversation between Liza and the learner; A comma-separated values (csv) file records the EEG signals of each participant recorded by the MindWave Mobile 2 device transferred to OpenVibe.

Since the MindWave Mobile 2 records EEG signals at a rate of 512 Hz but the attention values are calculated at a rate of 1 Hz, the csv file contained a great amount of redundant data of EEG signals. On average over all study participants, a session lasted about 39.35

minutes (2,361 seconds). That means that a csv file had about 1,208,832 lines of data entries. However, since the attention information was calculated by the eSense algorithm at each second, the redundant data in the csv file was thus eliminated to contain one data entry in each second interval. After that, the file was enriched with conversation data and time stamps such as when a new task was introduced or when feedback has been given.

The data analysis consists of multiple parts: 1) the changes of attention directly before and after feedback has been received; 2) the changes in attention over an entire task after receiving feedback; and 3) the difference between pre- and post-test questionnaire. The statistics method two-tailed *t*-test will be applied on these three analyses, since the hypothesis of this work is undirected, meaning attention can either decrease or increase after receiving feedback.

Changes in attention directly before and after receiving feedback. Feedback is given sixteen times during the entire conversation (one for each task solved). Consequently, it is necessary to determine how many data points will be evaluated for each time feedback is received. The following points were considered when thinking about the right amount of data points: when a task is introduced to a participant, the participant must first read and understand the question, and then solve it. After the user gives an answer, Liza either responds directly with the feedback, or asks the user to estimate their chances on having answered the task correctly. Both reading and thinking will affect the level of attention, overwriting any effect Liza's feedback might have. Therefore, for the direct comparison, a timeframe of 10 seconds before and after receiving feedback will be considered. With this timeframe, the mean of attention level, both before and the after receiving a feedback, was calculated and stored in two separate tables: one for positive feedback and one for negative feedback. Then, a paired *t*-test was conducted for each table. A paired *t*-test, or correlated *t*-test, is used when the mean difference in one subject is measured for a before and after comparison and when the two means are dependent (Xu et al., 2017).

Changes in attention over a task after feedback is received. The changes in attention during task-solving was also analyzed, to observe if a chatbot's feedback has an impact beyond the initial 10 seconds after receiving it. Q represents the moment Liza asks the question, and F the feedback after the question has been solved, with "$Q_n \rightarrow$" describing the entire task-solving process for the n_{th} task. Then, one comparison element will be from when feedback F_n is received up to when feedback F_{n+1} is received, with an entire task-solving of Q_{n+1} in between. For each *t*-test, the before and after will be compared:

$$\dots \Rightarrow \overbrace{Q_n \rightarrow}^{task\text{-}solving} \underbrace{F_n \Rightarrow Q_{n+1} \rightarrow}_{before} F_{n+1} \Rightarrow \underbrace{F_{n+1} \Rightarrow Q_{n+2} \rightarrow}_{after} F_{n+2} \Rightarrow \dots$$

The recording of each set consists of multiple seconds, from Liza introducing the task to the participant answering it. In almost all cases, the recordings of two consecutive sets are not the same length, and thus, will not have the same amount of data points. Therefore, an independent *t*-test will be used for each comparison.

Pre-test and post-test questionnaire. Both the pre-test and the post-test questionnaire provide a subjective rating for the learner's attention and can be used to compare with statistics data. The participants had to estimate their attention level for each question they solved with another Likert scale (very low = 1, low = 2, average = 3, high = 4, very high = 5). The actual level of attention, that was recorded using the MindWave Mobile 2 device during the task-solving, was also assigned with scores ("10 - 20" = 1, "20 - 40" = 2, "40 - 60" = 3, "60 - 80" = 4, "80 - 100" = 5). Then, a paired *t*-test was conducted, to determine if the self-estimated level of attention correlated with the recorded level of attention.

Results

During the recording of participant 10, there was a significant amount of time where no measurement took place (16 minutes in total). In the case of participant 6, there was a significant loss of information during 6 questions. Data of participant 10 and 6 were discarded. Participants 13, 14 and 15 also had corrupted data. Participants 13 and 14 each answered one question with "I do not know", resulting with Liza not giving feedback, but asking the participants if they would like to know the answer. Participant 15 had a significant time of no measurement in two questions. Only data regarding those questions were discarded, resulting in a total analysis of 16 recordings.

Changes in attention

10-second comparison. Here, both the results for the negative and positive feedback in the 10-second comparison are presented in two separated tables, the positive feedback in Table 2 and the negative feedback in Table 3. *n* indicates the amount of positive/negative feedback received, *x* is the mean of the 10 seconds before feedback is received and *y* is the mean of the 10 seconds after feedback is received. $x - y$ is the mean difference.

The *p*-value was not smaller than the significance level of 0.05 in any of the 16 recordings, indicating no significant difference in attention during the 10 seconds before, and 10 seconds after positive feedback is received. Thus, the H_0 hypothesis cannot be rejected. Table 3 shows the results of the paired *t*-test for incorrectly solved tasks, followed by negative feedback.

Similar to positive feedback, the H_0 hypothesis cannot be rejected in any of the 16 recordings, because none of the *t*-differences is significant at a significance level of 0.05.

Table 2 Results of paired *t*-test for the direct influence of positive feedback

ID	n	x	y	M(x)-M(y)	t-statistic	p-value
1	9	50.93	50.28	0.66	0.19	0.85
2	8	56.00	49.75	6.25	1.34	0.22
3	9	41.68	43.02	-1.34	-0.42	0.69
4	10	50.52	55.12	-4.60	-1.16	0.27
5	8	47.95	43.25	4.70	1.26	0.25
7	10	51.68	50.04	1.64	0.53	0.61
8	7	57.07	51.40	5.67	1.13	0.30
9	6	38.50	45.43	-6.92	-1.57	0.18
11	9	47.27	36.64	10.63	1.44	0.19
12	10	52.52	44.96	7.56	1.42	0.19
13	7	63.30	56.11	7.19	1.39	0.21
14	8	52.67	45.28	7.39	1.60	0.15
15	10	48.87	44.09	4.78	1.05	0.32
16	9	59.98	55.04	4.94	0.69	0.51
17	7	49.99	53.91	-3.92	-0.80	0.46
18	8	48.76	43.19	5.75	0.98	0.36

Table 3 Results of paired *t*-test for the direct influence of negative feedback

ID	n	x	y	M(x)-M(y)	t-statistic	p-value
1	7	51.33	45.69	5.66	1.39	0.21
2	8	48.39	46.79	1.60	0.10	0.93
3	7	41.73	43.26	-1.53	-0.33	0.75
4	6	45.93	53.18	-7.25	-1.66	0.16
5	8	50.61	51.46	-0.85	-0.32	0.76
7	6	58.05	56.15	1.90	1.07	0.33
8	9	46.04	46.71	-0.67	-0.16	0.87
9	10	44.02	40.41	3.61	0.88	0.40
11	7	48.93	39.89	9.04	2.36	0.06
12	6	64.37	57.38	6.99	1.16	0.30
13	8	41.45	43.01	-1.56	-0.27	0.80
14	7	43.13	35.93	7.20	1.36	0.22
15	5	54.26	54.60	-0.34	-0.08	0.94
16	7	51.23	52.35	-1.12	-0.30	0.78
17	9	46.09	52.69	-6.60	-1.34	0.22
18	8	48.37	41.56	6.81	1.20	0.27

H_0 hypothesis can be confirmed that neither positive nor negative feedback has a significant influence on human attention when comparing the recordings of 10 seconds.

Comparing changes in attention over the length of a task. Independent *t*-tests have been calculated for comparing the changes of attention over two tasks, after receiving positive feedback. Since the independent *t*-test was done for each task of each participant, it resulted in total 252 *t*-tests (see Appendix Table 5).

In total, the H_0 was rejected in 137 of all 252 cases (54.36%), meaning that in these cases, the *p*-value was smaller than 0.05. When looking at the type of feedback, positive feedback had a significant effect in 71 of these 137 cases (51.82%) and negative feedback had a significant effect in 66 of the 137 cases (48.18%). The H_1 hypothesis can be accepted in 88 cases of these 137 cases. When it comes to positive feedback, the predicted effect of the

H_1 hypothesis has been found in 52 cases. Respectively, the predicted effect of negative feedback has been found in 36 cases. However, all these instances do not occur evenly over the participants, with respect to the H_0 hypothesis. For participant 1, the H_0 hypothesis was rejected in 5 cases, for participant 13 in 6 cases, for participants 8 and 9 in 7 cases, for participants 11, 15, 16 and 17 in 8 cases, for participants 4, 5, and 7 in 9 cases, for participants 2, 3 and 14 in 10 cases, and lastly for participants 12 and 18 in 12 cases. The H_1 hypothesis reached a support rate in more than 50% percent in 15 participants, with 80% rate being the highest for participants 1, 2 and 3. Participant 4 could only support the H_1 hypothesis in 40% of the cases.

Pre- and post-questionnaire

Table 4 shows the results of the *t*-test for comparison between pre- and post-questionnaire. "Attention disorder" describes the paired *t*-test of the six pre- and post-questionnaire questions that are used to identify a possible attention deficit, whereas "self-report" shows the results of the self-estimated attention level compared to the recorded attention level.

For the first part, the "attention disorder", there was no indication of a significant difference at a significance level of 0.05. However, for the comparison of the self-report and recorded data, there were several participants (2, 3, 5, 8, 9 and 15) who indicated significant difference (i.e., *p*-values smaller than 0.05) between the self-reported and recorded attention levels. A deeper analysis for the self-reports of participants 1, 4, 7, 11, 12, 13, 14, 16, 17 and 18 who had non-significant difference between self-reported and

Table 4 Results of paired *t*-tests for pre- and post-questionnaire

ID	Attention disorder			Self-report		
	pre	post	*p*-value	self	recording	*p*-value
1	1.67	1.67	1.00	3.31	3.31	1.00
2	1.50	2.00	0.29	4.13	3.31	0.0005
3	2.67	2.00	0.17	3.75	2.38	1E-05
4	2.17	2.00	0.74	3.44	3.13	0.055
5	2.33	1.33	0.07	3.56	3.00	0.007
7	1.83	1.67	0.77	3.19	3.19	1.00
8	1.00	1.33	0.36	3.94	3.38	0.04
9	2.00	2.00	1.00	3.81	2.94	0.001
11	1.83	1.67	0.74	3.75	3.06	0.36
12	1.17	1.67	0.36	3.50	3.56	0.67
13	1.67	1.67	1.00	3.20	3.40	0.38
14	1.67	1.67	1.00	3.27	3.13	0.43
15	2.33	2.33	1.00	3.79	3.00	0.03
16	2.33	1.33	0.08	3.50	3.13	0.08
17	1.50	2.00	0.20	3.56	3.25	0.26
18	1.50	2.00	0.36	3.19	3.19	1.00

recorded attention levels was required. Of those participants, only participant 14 had a matching low self-reported and recorded attention level in task 12. However, the *p*-value of that task (and following that task) was higher than 0.05, and thus, the matching of low attention level was not significant.

Discussion and limitations

To observe the influence a chatbot's feedback has on human attention, an experiment was conducted, in which the human attention during a conversation with an educational chatbot was measured. In general, the results of the experiment were inconclusive and neither hypotheses can be fully accepted or rejected. When it comes to finding a reason for why a hypothesis might have been rejected, it is important to keep two main findings of current research in mind: sometimes, feedback simply has no effect and human attention has no fixed state, as it can fluctuate at any time (Bunce et al., 2010; Kluger & DeNisi, 1996).

The hypothesis H_0 can neither be fully rejected or accepted, since both the 10-second and task-comparison show contradicting results. There is no significant increase or decrease in attention after receiving any kind of feedback, thus supporting the claim of the H_0 hypothesis, that neither positive nor negative feedback has an influence on human attention. There are a few possible reasons that could explain the absence of an effect. In the case of negative feedback, when a task has been solved falsely, Liza offers the participant the correct solution. If the participant accepts, they will read the solution in those 10 seconds, otherwise Liza introduces a new task. Except for one case, participants always accepted the offer of a solution. Therefore, negative feedback might not have had an impact in the 10 seconds, since the participants were busy reading the correct solution, before solving a new task. In the case of positive feedback, Liza goes straight to a new task. The lack of influence could be explained by the fact that the participant did not have enough time to process the feedback, since a new task of a different reasoning topic was immediately introduced. Another possible reason for a lack of effect of both positive and negative might be that the time frame of 10 seconds was simply too small to see a difference, which becomes apparent when comparing the attention over two tasks. In the task-comparison, the H_0 hypothesis is rejected multiple times, roughly in half of all comparisons. These results are consistent with the research regarding the effect of feedback, being that feedback can be both influential (e.g., Arbel et al. (2020); Hattie & Timperley (2007)) and non-influential (Kluger & DeNisi, 1996). The simplest reason for this phenomenon is that the feedback of a chatbot does not have an equal influence on the attention of human, as humans do not react untimely to stimuli. This reason is further supported by the results of this work. Participant 1, for example, had a significant reaction to 5 of the 16 pieces of feedback received, whereas participant 12 had a reaction to 12 of the 16 times of feedback.

Regarding the H_1 hypothesis, the study shows no conclusive result regarding the rejection or acceptance of it. In 64% of the comparisons, the predicted effect occurred, meaning that positive feedback was followed by a decrease of attention, and negative feedback by an increase of attention. The predicted influence of positive feedback is more dominant, with a decrease in attention occurring in 73.24% of the times, when a participant has received positive feedback. This shows, that the positive feedback of an educational chatbot can decrease the attention of a human, and should therefore only be used when necessary. Nonetheless, the H_1 hypothesis could not be accepted in every case. This might be due to the variance of difficulty across the reasoning topics. The Base Rate Fallacy was declared the most difficult one, and participants often did not want to solve these kinds of tasks. Receiving negative feedback in these situations, where the participant already expected to fail will not have the same influence as when the participant is unsure whether they have solved the task correctly. Consequently, receiving positive feedback in those situations could lead to an increase of attention, as the participant did not expect this, and therefore wants to pay more attention to the following tasks. In contrast, the Sunk Cost Fallacy has been found to be the easiest reasoning topic. These tasks use "either/or" questions. If the participant answered the first Sunk Cost Fallacy question falsely, they easily worked out the answer for the next task. Here, negative feedback might increase the attention, as the participant would have expected their answer to be incorrect. However, this instance works in favor of the H_1 hypothesis.

This study had several limitations and thus, the results of this study require a careful consideration. First, due to the small number of participants, it was not possible to implement a control group. All participants received feedback, and while there has been a significant influence of feedback in some cases, it is not sufficient to say that this effect only comes from the feedback of a chatbot. Second, in some cases, Liza was not able to understand the answer of the participant. Furthermore, Liza interpreted the input of participant sometimes falsely, e.g., when the participant asked for help while solving a task, Liza interpreted "help" as an answer to the task. Third, many participants noted that Liza did not give enough time to solve the task. When solving a task, while thinking about a solution, Liza is waiting for the learner's solution. If one minute passes without a solution, there is no input from the participant, since they are thinking of the correct solution. Liza interprets in the situation of no learners' and letting them know that they can take as much time as they need and offering help. She does this multiple times, until the participant answers. As commented by one participant, this led to frustration, and a disruption of the thinking process. These technical issues require to improve Liza' ability to interpret user's answers more accurately.

The recorded and analyzed data relies solely on the accuracy of NeuroSky's algorithm that is a black-box and thus, we do not know how the algorithm was developed. Therefore,

a concurrent validation of this algorithm may help to compare results provided by the NeuroSky's algorithm.

Conclusion and future work

This paper proposed a study to answer the research question "How can a chatbot's feedback influence human attention?". In the state of the art, feedback was found to overall improve learning performance. To analyze the relationship of positive and negative feedback to learner's attention, an experiment was carried out. 18 study participants had a conversation with an educational chatbot for improving human reasoning, while their attention was constantly measured with a mobile EEG device.

This study results in three contributions. First, it was found that there was a significant attention effect occurring in 54% of the times, after the educational chatbot showed feedback and the attention measurement took place over the length of a task. Second, when differentiating the type of feedback, positive feedback had a significant effect in 71 of these 137 cases (51.82%) and negative feedback had a significant effect in 66 of the 137 cases (48.18%). Third, statistical results showed that there was no significant difference in attention at the significant level of 0.05 during the 10 seconds before, and 10 seconds after positive feedback is received. Similar is the case for negative feedback. Results of this study recommend us to measure attention over a longer period.

While this research has produced results which are inconclusive, that does not mean that the question of how educational chatbots affect human attention is unanswerable. The results of this paper points in two directions, which are both equally acceptable: 1) feedback of an educational chatbot has no effect; 2) feedback of an educational chatbot can influence human attention, with positive feedback being a slightly larger driver. Despite the lack of a control group, and the inconclusive results, this paper was able to provide a first step into the understanding of how educational chatbots can influence human attention.

If an educational chatbot is used to ease the workload of a teacher, developers must accept that feedback will not have the desired effect on every student equally. Until the research question can be fully answered, it is recommended that an educational chatbot should not solely rely on feedback as a tool to control the attention of a student.

Based on this work, further questions have been raised: Is the feedback of an educational chatbot really the cause of the change, or is there just a correlation? Are there other methods a chatbot can use to influence human attention, that do not involve the constant measurement of physiological data?

Education is an important factor in everyone's life. It is crucial that all teachers, human or computer, take the role of teaching seriously and know about the effects of their teaching tools. Although this work was not able to find a clear answer to the proposed research question, future research in this field is still necessary.

Appendix: Results of the task comparison

The following table contains the entirety of results of the task comparison. The column "Means of Feedback" contains the mean of the before and after, as described in the data analysis in Section "Data Collection and Analysis Methods". "Type" stands for the type of feedback, either positive (p) or negative(n). "Results of the *t*-test" contains the *t*-statistic, degree of freedom (*df*) and the *p*-value. If H_0 is marked with an "X", the H_0 hypothesis was rejected for this feedback; if H_1 is marked with an "X", the H_1 hypothesis was accepted. F_n stands for the nth feedback given, with n = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16}. If an entire row is marked with a minus (-), there was not enough data to be analyzed.

Table 5 Complete results of task comparison

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
1	F_1	47.83	52.07	p	-1.61	204	0.11		
	F_2	52.07	63.70	p	-4.35	196	2.22E-05	X	X
	F_3	63.70	48.15	p	5.79	210	2.46E-08	X	X
	F_4	48.15	52.78	n	-1.93	158	0.056		
	F_5	52.78	61.90	p	-2.76	60	0.007	X	
	F_6	61.90	62.99	p	-0.32	66	0.75		
	F_7	62.99	59.58	n	1.34	144	0.18		
	F_8	59.58	57.29	n	0.91	136	0.36		
	F_9	57.29	64.29	n	-2.53	118	0.01	X	X
	F_{10}	64.29	65.48	p	-0.33	99	0.74		
	F_{11}	65.48	58.19	p	2.07	92	0.04	X	X
	F_{12}	58.19	52.63	n	1.75	111	0.08		
	F_{13}	52.63	47.00	p	1.82	85	0.07		
	F_{14}	47.00	50.18	p	-1.27	56	0.21		
	F_{15}	50.18	47.25	p	1.24	97	0.21		
	F_{16}	47.25	47.26	n	-0.001	49	0.99		
2	F_1	64.58	60.23	p	2.16	159	0.03	X	X
	F_2	60.23	67.32	p	-3.30	179	0.001	X	
	F_3	67.32	62.09	p	2.72	213	0.007	X	X
	F_4	62.09	61.25	n	0.52	268	0.60		
	F_5	61.25	51.34	p	6.01	261	6.06E-09	X	X
	F_6	51.34	59.11	n	-3.92	165	0.0001	X	X
	F_7	59.11	52.27	p	3.70	149	0.0003	X	X
	F_8	52.27	53.84	n	-0.0821	174	0.41		
	F_9	53.84	51.91	p	0.77	111	0.44		
	F_{10}	51.91	46.57	p	1.97	106	0.051		
	F_{11}	46.57	49.98	n	-1.54	98	0.13		
	F_{12}	49.98	58.30	n	-4.82	196	2.82E-06	X	X
	F_{13}	58.30	44.40	n	6.30	160	2.74E-09	X	X
	F_{14}	44.40	40.05	p	1.71	128	0.08		
	F_{15}	40.05	44.92	n	-2.09	122	0.03	X	X
	F_{16}	44.92	23.22	n	6.19	11	6.77E-05	X	

Table 5 (Cont'd.)

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
3	F_1	54.31	36.69	p	4.81	126	4.26E-06	X	X
	F_2	36.69	37.26	p	-0.18	113	0.86		
	F_3	37.26	29.41	p	3.56	125	0.0005	X	X
	F_4	29.41	34.06	n	-2.86	159	0.004	X	X
	F_5	34.06	38.97	p	-2.67	150	0.008	X	
	F_6	38.97	50.62	n	-6.04	134	1.40E-08	X	X
	F_7	50.62	37.53	p	6.10	128	1.14E-08	X	X
	F_8	37.53	33.84	n	1.59	126	0.11		
	F_9	33.84	33.51	p	0.16	114	0.87		
	F_{10}	33.51	35.52	n	-1.01	146	0.31		
	F_{11}	35.52	44.31	n	-4.44	120	2.02E-05	X	X
	F_{12}	44.31	42.70	n	0.70	90	0.49		
	F_{13}	42.70	45.41	p	-0.99	82	0.33		
	F_{14}	45.41	36.39	p	3.57	68	0.0006	X	X
	F_{15}	36.39	45.07	p	-4.09	84	9.54E-05	X	X
	F_{16}	45.07	39.17	n	2.04	44	0.04	X	
4	F_1	63.76	59.73	p	1.60	125	0.11		
	F_2	59.73	63.71	p	-1.96	117	0.05		
	F_3	63.71	59.89	p	2.68	261	0.007	X	X
	F_4	59.89	46.51	n	8.50	143	2.26E-14	X	
	F_5	46.51	53.69	p	-3.79	73	0.0003	X	
	F_6	53.69	46.98	p	2.91	77	0.004	X	X
	F_7	46.98	47.28	p	-0.15	76	0.88		
	F_8	47.28	53.17	n	-2.55	70	0.012	X	X
	F_9	53.17	46.25	p	2.54	98	0.012	X	
	F_{10}	46.25	51.56	p	-1.62	59	0.89		
	F_{11}	51.56	51.16	n	0.13	44	0.89		
	F_{12}	51.16	55.56	n	-2.84	336	0.004	X	X
	F_{13}	55.56	50.16	p	2.69	109	0.008	X	X
	F_{14}	50.16	51.72	p	-0.72	80	0.47		
	F_{15}	51.72	51.12	n	0.33	71	0.74		
	F_{16}	51.12	60.08	n	-5.23	9	0.0005	X	X

Table 5 (Cont'd.)

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
5	F_1	55.49	56.24	p	-0.39	155	0.69		
	F_2	56.24	51.78	n	2.96	268	0.003	X	
	F_3	51.78	51.61	n	0.12	271	0.90		
	F_4	51.61	49.66	n	1.08	191	0.28		
	F_5	49.66	51.54	p	-0.98	210	0.38		
	F_6	51.54	43.82	p	4.22	178	3.89E-05	X	X
	F_7	43.82	42.07	p	0.81	170	0.42		
	F_8	42.07	53.99	n	-5.47	178	1.49E-07	X	X
	F_9	53.99	46.53	p	2.89	96	0.004	X	X
	F_{10}	46.53	52.40	p	-1.86	116	0.06		
	F_{11}	52.40	46.35	n	2.44	86	0.01	X	
	F_{12}	46.35	57.90	n	-6.59	174	5.11E-10	X	X
	F_{13}	57.90	52.84	p	2.05	115	0.04	X	X
	F_{14}	52.84	54.36	n	-0.64	106	0.52		
	F_{15}	54.36	50.09	p	2.39	259	0.01	X	X
	F_{16}	50.09	57.11	n	-2.74	51	0.008	X	X
7	F_1	69.92	72.38	p	-1.00	118	0.32		
	F_2	72.38	60.42	n	7.80	221	2.42E-13	X	
	F_3	60.42	55.42	p	4.18	632	3.27E-05	X	X
	F_4	55.42	57.66	n	-1.25	113	0.21		
	F_5	57.66	58.81	p	-0.49	78	0.62		
	F_6	58.81	59.27	p	-0.19	82	0.85		
	F_7	59.27	43.81	p	7.00	138	9.91E-11	X	X
	F_8	43.81	54.58	n	-4.81	152	3.57E-06	X	X
	F_9	54.58	51.13	p	1.53	123	0.13		
	F_{10}	51.13	54.69	n	-1.72	125	0.08		
	F_{11}	54.69	47.11	p	4.04	72	8.83E-05	X	X
	F_{12}	47.11	51.55	n	-2.43	168	0.01	X	X
	F_{13}	51.55	40.35	p	4.14	136	6.03E-05	X	X
	F_{14}	40.35	42.70	p	-0.79	118	0.43		
	F_{15}	42.70	54.83	p	-4.46	107	2E-05	X	
	F_{16}	54.83	61.93	n	-2.31	66	0.02	X	X

Table 5 (Cont'd.)

ID		Means of Feedback			Results of t-test				
		before	after	type	t-statistic	df	p-value	H_0	H_1
8	F_1	66.56	63.35	n	1.53	111	0.13		
	F_2	63.35	61.23	p	1.11	91	0.27		
	F_3	61.23	64.03	p	-1.76	275	0.08		
	F_4	64.03	62.49	n	0.71	195	0.48		
	F_5	62.49	55.32	p	2.32	107	0.02	X	X
	F_6	55.32	53.84	n	0.51	89	0.61		
	F_7	53.84	51.71	n	1.30	105	0.20		
	F_8	51.71	53.21	n	-0.93	217	0.35		
	F_9	53.21	48.22	p	1.93	96	0.05		
	F_{10}	48.22	57.10	n	-3.29	103	0.001	X	
	F_{11}	57.10	50.11	p	3.92	125	0.0001	X	X
	F_{12}	50.11	53.32	n	-1.66	92	0.10		
	F_{13}	53.32	65.26	n	-4.30	135	3.26E-05	X	X
	F_{14}	65.26	57.30	n	3.26	118	0.001	X	X
	F_{15}	57.30	42.82	p	9.84	353	2.31E-20	X	
	F_{16}	42.82	61.90	n	-4.94	33	2.22E-05	X	X
9	F_1	51.42	48.48	n	1.71	240	0.09		
	F_2	48.48	44.47	p	2.44	291	0.02	X	X
	F_3	44.47	42.83	p	1.12	385	0.27		
	F_4	42.83	45.65	n	-1.87	280	0.06		
	F_5	45.65	41.96	p	2.40	262	0.02	X	X
	F_6	41.96	37.87	n	2.05	106	0.04	X	
	F_7	37.87	49.10	n	-5.67	104	1.26E-07	X	X
	F_8	49.10	50.60	n	-0.96	377	0.34		
	F_9	50.60	48.54	p	1.19	229	0.24		
	F_{10}	48.54	42.22	n	3.78	201	0.0002	X	
	F_{11}	42.22	40.51	p	1.00	270	0.32		
	F_{12}	40.51	50.52	n	-5.68	242	3.89E-08	X	X
	F_{13}	50.52	54.03	n	-1.66	119	0.10		
	F_{14}	54.03	53.50	n	0.23	139	0.82		
	F_{15}	53.50	57.23	p	-1.95	191	0.05		
	F_{16}	57.23	49.53	n	3.47	51	0.001	X	

Table 5 (Cont'd.)

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
10	F_1	61.45	50.42	p	4.65	137	7.76E-06	X	X
	F_2	50.42	50.08	n	0.16	172	0.87		
	F_3	50.08	50.91	n	-0.36	211	0.72		
	F_4	50.91	55.20	n	-1.91	182	0.06		
	F_5	55.20	44.13	p	5.76	192	3.26E-08	X	X
	F_6	44.13	45.38	n	-0.61	179	0.54		
	F_7	45.38	56.33	p	-2.78	59	0.007	X	
	F_8	56.33	53.47	n	0.73	59	0.47		
	F_9	53.47	43.09	p	4.33	216	2.33E-05	X	X
	F_{10}	43.09	46.86	p	-1.35	114	0.18		
	F_{11}	46.86	51.37	p	-1.68	97	0.10		
	F_{12}	51.37	42.05	n	4.44	227	1.39E-05	X	
	F_{13}	42.05	47.90	p	-2.20	148	0.03	X	
	F_{14}	47.90	51.04	p	-1.18	147	0.24		
	F_{15}	51.04	46.25	p	2.44	257	0.02	X	X
	F_{16}	46.25	36.55	n	1.68	41	0.003	X	
12	F_1	67.40	69.02	p	-0.51	159	0.61		
	F_2	69.02	76.70	p	-2.67	169	0.008	X	
	F_3	76.70	65.96	p	3.56	226	0.0005	X	X
	F_4	65.96	52.16	n	4.44	257	1.32E-05	X	
	F_5	52.16	49.59	p	0.88	247	0.38		
	F_6	49.59	44.26	p	2.14	199	0.03	X	X
	F_7	44.26	60.80	p	-5.08	95	1.85E-06	X	
	F_8	60.80	67.70	n	-1.83	130	0.07		
	F_9	67.70	53.38	p	5.14	143	8.86E-07	X	X
	F_{10}	53.38	48.17	n	2.18	182	0.03	X	
	F_{11}	48.17	52.75	p	-1.83	141	0.07		
	F_{12}	52.75	64.47	n	-3.41	134	0.001	X	X
	F_{13}	64.47	56.74	p	2.23	147	0.03	X	X
	F_{14}	56.74	62.99	n	-2.65	279	0.008	X	X
	F_{15}	62.99	68.16	p	-2.24	253	0.03	X	
	F_{16}	68.16	53.53	n	6.62	123	9.84E-10	X	

Table 5 (Cont'd.)

ID		Means of Feedback			Results of t-test				
		before	after	type	t-statistic	df	p-value	H_0	H_1
13	F_1	62.09	68.04	n	-2.77	151	0.006	X	X
	F_2	68.04	67.89	p	0.07	149	0.95		
	F_3	67.89	59.83	p	3.51	224	0.0005	X	X
	F_4	59.83	62.22	n	-0.89	133	0.37		
	F_5	62.22	58.58	p	1.21	65	0.23		
	F_6	58.58	67.40	n	-3.02	63	0.004	X	X
	F_7	67.40	62.26	p	1.90	150	0.06		
	F_8	62.26	54.74	n	2.54	126	0.01	X	
	F_9	54.74	50.48	p	1.16	96	0.25		
	F_{10}	50.48	54.60	n	-1.18	93	0.24		
	F_{11}	54.60	58.03	p	-1.35	160	0.18		
	F_{12}	58.03	56.24	n	0.76	147	0.45		
	F_{13}	56.24	48.86	p	2.73	140	0.007	X	X
	F_{14}	48.86	51.53	n	-1.16	110	0.25		
	F_{15}	-	-	-	-	-	-	-	-
	F_{16}	45.90	32.58	n	3.74	47	0.0005	X	
14	F_1	74.99	49.58	p	9.09	170	2.51E-16	X	X
	F_2	49.58	61.53	p	-5.10	175	8.58E-07	X	
	F_3	61.53	54.36	p	3.31	329	0.001	X	X
	F_4	54.36	55.80	n	-0.61	192	0.54		
	F_5	55.80	56.32	p	-0.09	25	0.93		
	F_6	56.32	73.49	p	-2.87	27	0.008	X	X
	F_7	73.49	52.14	p	7.98	152	3.39E-13	X	X
	F_8	52.14	58.09	n	-2.32	200	0.02	X	X
	F_9	58.09	58.91	p	-0.31	152	0.76		
	F_{10}	58.91	47.07	n	5.33	120	4.67E-07	X	
	F_{11}	47.07	38.88	n	4.72	296	3.63E-06	X	
	F_{12}	38.88	42.44	n	-1.92	205	0.06		
	F_{13}	42.44	44.56	p	-0.93	99	0.36		
	F_{14}	44.56	36.86	n	3.85	68	0.0003	X	
	F_{15}	-	-	-	-	-	-	-	-
	F_{16}	43.08	34.82	n	2.10	32	0.04	X	

Table 5 (Cont'd.)

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
15	F_1	43.38	41.49	p	0.82	126	0.41		
	F_2	41.49	45.11	p	-2.14	220	0.03	X	
	F_3	45.11	39.24	p	3.32	286	0.001	X	X
	F_4	39.24	48.25	n	-5.08	283	6.80E-07	X	X
	F_5	48.25	55.46	p	-3.04	92	0.003	X	
	F_6	55.46	59.74	p	-1.47	105	0.14		
	F_7	59.74	43.38	p	6.63	96	1.94E-09	X	X
	F_8	43.38	57.14	n	-6.30	170	2.46E-09	X	X
	F_9	57.14	58.37	p	-0.50	122	0.62		
	F_{10}	58.37	53.20	p	2.36	99	0.02	X	X
	F_{11}	53.20	51.06	n	1.44	184	0.15		
	F_{12}	51.06	59.51	n	-4.98	126	2.03E-06	X	X
	F_{13}	59.51	61.13	p	-0.50	57	0.62		
	F_{14}	61.13	60.95	n	0.06	53	0.96		
	F_{15}	-	-	-	-	-	-	-	-
	F_{16}	-	-	-	-	-	-	-	-
16	F_1	62.32	50.09	p	4.93	161	2.04E-06	X	X
	F_2	50.09	59.68	n	-4.73	208	4.12E-06	X	X
	F_3	59.68	60.68	p	-0.52	336	0.60		
	F_4	60.68	52.94	n	3.76	255	0.0002	X	
	F_5	52.94	51.85	n	0.40	103	0.69		
	F_6	51.85	58.35	p	-1.75	103	0.08		
	F_7	58.35	53.51	p	1.45	89	0.15		
	F_8	53.51	46.33	n	2.47	145	0.01	X	
	F_9	46.33	51.09	p	-1.35	61	0.18		
	F_{10}	51.09	57.76	p	-1.62	64	0.11		
	F_{11}	57.76	49.88	p	2.25	61	0.03	X	X
	F_{12}	49.88	57.81	n	-3.44	189	0.0007	X	X
	F_{13}	57.81	56.85	p	0.33	135	0.74		
	F_{14}	56.85	56.66	p	0.06	97	0.95		
	F_{15}	56.66	44.83	n	4.91	48	1.09E-05	X	
	F_{16}	44.83	64.34	n	-6.72	32	1.38E-07	X	X

Table 5 (Cont'd.)

ID		Means of Feedback			Results of *t*-test				
		before	after	type	*t*-statistic	*df*	*p*-value	H_0	H_1
17	F_1	55.70	55.37	p	0.13	130	0.90		
	F_2	55.37	59.75	n	-2.23	153	0.03	X	X
	F_3	59.75	53.64	p	3.45	225	0.0007	X	X
	F_4	53.64	61.47	n	-4.01	168	9.07E-05	X	X
	F_5	61.47	62.92	p	-0.61	92	0.55		
	F_6	62.92	62.68	p	0.08	92	0.94		
	F_7	62.68	50.84	n	4.81	73	7.95E-06	X	
	F_8	50.84	48.73	n	0.96	111	0.34		
	F_9	48.73	47.04	n	0.60	134	0.55		
	F_{10}	47.04	52.22	n	-1.88	134	0.06		
	F_{11}	52.22	58.89	p	-2.10	105	0.04	X	
	F_{12}	58.89	50.97	n	2.34	118	0.02	X	
	F_{13}	50.97	40.74	p	2.80	46	0.007	X	X
	F_{14}	40.74	61.76	p	-4.90	58	7.99E-06	X	
	F_{15}	61.76	55.36	n	1.60	91	0.11		
	F_{16}	55.36	49.00	n	1.30	8	0.23		
18	F_1	74.41	67.97	p	2.79	173	0.006	X	X
	F_2	67.97	62.02	n	2.68	218	0.008	X	
	F_3	62.02	55.40	p	3.20	201	0.002	X	X
	F_4	55.40	59.66	p	-1.32	85	0.19		
	F_5	59.66	58.02	n	0.38	92	0.71		
	F_6	58.02	41.79	n	4.61	60	2.17E-05	X	
	F_7	41.79	48.33	n	-2.92	256	0.004	X	X
	F_8	48.33	59.68	n	-4.45	164	1.56E-05	X	X
	F_9	59.68	52.92	p	1.86	87	0.07		
	F_{10}	52.92	66.48	p	-3.84	82	0.0002	X	
	F_{11}	66.48	55.17	p	3.47	81	0.0008	X	X
	F_{12}	55.17	46.30	n	2.53	89	0.01	X	
	F_{13}	46.30	35.70	p	3.09	64	0.003	X	X
	F_{14}	35.70	51.38	p	-4.06	67	0.0001	X	
	F_{15}	51.38	42.94	n	2.49	75	0.01	X	
	F_{16}	42.94	44.07	n	-0.27	43	0.79		

Abbreviations

ADHD: Attention deficit hyperactivity disorder; CSV: Comma-separated values; EEG: Electroencephalogram; SVM: Support Vector Machine.

Endnotes

[1] https://neurosky.com/biosensors/eeg-sensor/algorithms (Accessed: 10/05/2023)
[2] https://physiological-edu.informatik.hu-berlin.de/ (Accessed: 10/05/2023)
[3] https://rasa.com/ (Accessed: 10/05/2023)
[4] https://neurosky.com/biosensors/eeg-sensor/biosensors/ (Accessed: 10/05/2023)
[5] http://openvibe.inria.fr/ (Accessed: 10/05/2023)

Author details
[1] Humboldt-Universität zu Berlin, Germany
[2] Charité – Universitätsmedizin Berlin, Germany

References

Aftanas, L. I., & Golocheikine, S. A. (2001). Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: High-resolution EEG investigation of meditation. *Neuroscience Letters*, *310*, 57–60.

Al'Omari, T., & Balushi, H. A. (2015). The influence of paying attention in classroom on students' academic achievement in terms of their comprehension and recall ability. In *Proceedings of INTCESS15- 2nd International Conference on Education and Social Sciences* (pp. 684–693). OCERINT.

Arbel, Y., Feeley, E., & He, X. (2020). The effect of feedback on attention allocation in category learning: An eye tracking study. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.559334

Bagherzadeh, Y., Baldauf, D., Pantazis, D., & Desimone, R. (2020). Alpha synchrony and the neurofeedback control of spatial attention. *Neuron*, *105*(3), 577–587. https://doi.org/10.1016/j.neuron.2019.11.001

Battle, D. E. (2013). Diagnostic and statistical manual of mental disorders (dsm). *CoDAS*, *25*(2), 191–192. https://doi.org/10.1590/s2317-17822013000200017

Bitner, R. A., & Le, N.-T. (2021). Can EEG-devices differentiate attention values between incorrect and correct solutions for problem-solving tasks? *Journal of Information and Telecommunication*, *6*(2), 121–140. https://doi.org/10.1080/24751839.2021.1950319

Brackbill, Y., & O'Hara, J. (1958). The relative effectiveness of reward and punishment for discrimination learning in children. *Journal of Comparative and Physiological Psychology*, *51*(6), 747–751. https://doi.org/10.1037/h0038282

Bunce, D. M., Flens, E. A., & Neiles, K. Y. (2010). How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education*, *87*(12), 1438–1443. https://doi.org/10.1021/ed100409p

Chiang, H. S., Hsiao, K. L., & Liu, L. C. (2018). EEG-based detection model for evaluating and improving learning attention. *Journal of Medical and Biological Engineering*, *38*, 847–856. https://doi.org/10.1007/s40846-017-0344-z

Delgado, P., Vargas, C., Ackerman, R., & Salmeron, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, *25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/10.1016/j.copsyc.2019.03.005

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

James, W. (1890). *The principles of psychology, vol I*. Henry Holt and Co.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, *37*, 43–75. https://doi.org/10.1146/annurev.ps.37.020186.000355

Junco, R., & Cotten, S. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers and Education*, *59*(2), 505–514. https://doi.org/10.1016/j.compedu.2011.12.023

Keller, A. S., Davidesco, I., & Tanner, K. D. (2020). Attention matters: How orchestrating attention may relate to classroom learning. *CBE—Life Sciences Education*, *19*(3), 19:fe5, 1–9. https://doi.org/10.1187/cbe.20-05-0106

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470752937.ch16

Le, N.-T., & Wartschinski, L. (2018). A Cognitive Assistant for improving human reasoning skills. *International Journal of Human-Computer Studies*, *117*, 45–54. https://doi.org/10.1016/j.ijhcs.2018.02.005

Liaw, S.-S., Huang, H.-M., & Chen, G.-D. (2007). Surveying instructor and learner attitudes toward e-learning. *Computers & Education*, *49*(4), 1066–1080. https://doi.org/10.1016/j.compedu.2006.01.001

Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Frontiers in Computational Neuroscience*, *14*. https://doi.org/10.3389/fncom.2020.00029

Linden, M., Habib, T., & Radojevic, V. (1996). A controlled study of the effects of EEG biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities. *Biofeedback and Self-Regulation*, *21*(1), 35–49. https://doi.org/10.1007/bf02214148

Liu, N.-H., Chiang, C.-Y., & Chu, H.-C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors*, *13*(8), 10273–10286. https://doi.org/10.3390/s130810273

Lodge, J. M., & Harrison, W. J. (2019). The role of attention in learning in the digital age. *Journal of Biology and Medicine*, *92*, 21–28.

Lufi, D., & Haimov, I. (2018). Effects of age on attention level: Changes in performance between the ages of 12 and 90. *Aging, Neuropsychology, and Cognition*, *26*(6), 904–919. https://doi.org/10.1080/13825585.2018.1546820

Mancas, M., & Ferrera, V. P. (2016). How to measure attention? In M. Mancas, V. P. Ferrera, N. Riche & J. G. Taylor (Eds.), *From Human Attention to Computational Attention* (pp. 21–38). Springer. https://doi.org/10.1007/978-1-4939-3435-5_3

McDowd, J. M. (2007). An overview of attention: Behavior and brain. *Journal of Neurologic Physical Therapy*, *31*(3), 98–103. https://doi.org/10.1097/npt.0b013e31814d7874

McDowd, J. M., Vercruyssen, M., & Birren, J. E. (1991). Aging, divided attention, and dual-task performance. In D. Damos (Ed.), *Multiple-task performance* (pp. 387–414). CRC Press.

Meyer, W. J., & Offenbach, S. I. (1962). Effectiveness of reward and punishment as a function of task complexity. *Journal of Comparative and Physiological Psychology*, *55*(4), 532–534. https://doi.org/10.1037/h0049119

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Ni, D., Wang, S., & Liu, G. (2020). The EEG-based attention analysis in multimedia m-learning. *Computational and Mathematical Methods in Medicine*, *2020*, 4837291. https://doi.org/10.1155/2020/4837291

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: Electrophysiological signals predict errors up to 20 s before they occur. *Journal of Neuroscience*, *29*(26), 8604–8611. https://doi.org/10.1523/jneurosci.5967-08.2009

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clinical Neurophysiology*, *117*(9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Pavlygina, R. A., Davydov, V. I., Sakharov, D. S., Tutushkina, M. V., & Pryamonosova, A. A. (2011). The EEG during solution of mathematical logical tasks. *Neuroscience and Behavioral Physiology*, *42*(1), 28–35. https://doi.org/10.1007/s11055-011-9528-0

Ray, W. J., & Cole, H. W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, *228*(4700), 750–752. https://doi.org/10.1126/science.3992243

Schumann, F., Steinborn, M. B., Kürten, J., Cao, L., Händel, B. F., & Huestegge, L. (2022). Restoration of attention by rest in a multitasking world: Theory, methodology, and empirical evidence. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.867978

Sezer, A., İnel, Y., Seçkin, A. Ç., & Uluçınar, U. (2015). An investigation of university students' attention levels in real classroom settings with NeuroSky's MindWave mobile (EEG) device. In *Proceedings of International Educational Technology Conference* (pp. 88–101). International Educational Technology Conference.

Sohlberg, M. M., & Mateer, C. A. (1987). Effectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology*, *9*(2), 117–130. https://doi.org/10.1080/01688638708405352

Teplan, M. (2002). Fundamental of EEG measurement. *Measurement Science Review*, *2*, 2.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*(1/4), 212–222. https://doi.org/10.2307/1415413

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Wartschinski, L., Le, N.-T., & Pinkwart, N. (2017). A conversational agent for the improvement of human reasoning skills. *Bildungsräume 2017 Gesellschaft für Informatik*, 249–260.

Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M., & Feng, C. (2017). The differences and similarities between two-sample t-test and paired t-test. *Shanghai Archives of Psychiatry*, *29*, 184–188. https://doi.org/10.11919/j.issn.1002-0829.217070

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *Journal on Image and Video Processing*, *2017*, 80. https://doi.org/10.1186/s13640-017-0228-8

Zanto, T. P., Pan, P., Liu, H., Bollinger, J., Nobre, A. C., & Gazzaley, A. (2011). Age-related changes in orienting attention in time. *Journal of Neuroscience*, *31*(35), 12461–12470. https://doi.org/10.1523/jneurosci.1149-11.2011

Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research and Practice in Technology Enhanced Learning (RPTEL)
is an open-access journal and free of publication fee.