Journal on Multimodal User Interfaces (2024) 18:21–41
https://doi.org/10.1007/s12193-023-00418-5

ORIGINAL PAPER

A social robot as your reading companion: exploring the relationships between gaze patterns and knowledge gains

Xuan Liu[1] · Jiachen Ma[2] · Qiang Wang[1]

Received: 18 October 2022 / Accepted: 22 September 2023 / Published online: 12 October 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract
Intelligent tutoring systems have been widely used in educational activities over the past 20 years. With significantly less effort put into writing or reading assistance, the majority of intelligent tutoring systems focus on mathematics or problem solving activities. However, with the development of E-reading-centered E-education, how to improve students' learning performance during reading has become increasingly important. Therefore, in this paper, we take a first step in the direction of an adaptive intelligent tutoring system by investigating how different reading strategies relate to knowledge gain based on gaze features and how an embodied social robot affects gaze patterns and reading strategies. The findings showed that different knowledge gains have significant differences in scanning methods and reading depth, and that the feedback given by social robots significantly affects participants' gaze patterns during the whole reading process. To automatically differentiate between two levels of knowledge gain, several prediction experiments based on various reading strategy-related gaze features were carried out. The results demonstrate that saccades are the best predictors of knowledge gain, with the best model having an average accuracy of 74.2%. Finally, real-time simulation experiments were conducted with sixty participants using the leave-one-out method, and an accurate prediction of the level of knowledge gain of 71. 5% was achieved.

Keywords Eye-gaze · Intelligent tutoring systems · Reading support · Machine learning

1 Introduction

Reading is a fundamental skill that supports people in the learning process, but it is not a straightforward task that's easy to master. Reading is a complex process that draws on many different skills. Together, these skills lead to the ultimate goal of reading: reading comprehension or knowledge-gain. Teaching reading skills begins at a young age, when students learn the construction of words from letters, and continues through reading comprehension (i.e., making sense of a text). However, even after students become fluent readers, they may struggle with gaining knowledge from texts, often forgetting what they have just read. Although there has been extensive research on the support of beginning reading skills [13] and reading comprehension [21], little research has investigated how to support fluent readers in optimizing their knowledge gain. In this paper, we aim to automatically analyze the reading patterns, operationalized through gaze data, of participants as a first step toward developing automated feedback that can support students in optimizing their knowledge gain.

Reading patterns analysis, mainly conducted by gaze data computation, aims to evaluate people's mental states or reading strategies. In previous studies [11, 18, 19, 32, 46], researchers measured participants' attention level variation during reading based on their gaze data and using self-report or thought probes as ground truth. Although, detecting and interrupting mind wander sounds promising in optimizing reading performance, this method is limited since it relies on the accuracy and honesty of participants' mind-wandering reports, and the measurement interrupts the natural flow of the reading activity. Measurement reading strategies are an

✉ Qiang Wang
wangqiang@hit.edu.cn

Xuan Liu
hitliuxuan@163.com

Jiachen Ma
hitmjc@163.com

[1] Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, People's Republic of China

[2] Department of Control Science and Engineering, Harbin Institute of Technology, Weihai 150001, People's Republic of China

alternative without the aforementioned limits. Appropriate reading strategies could help the learner focus on the most important passages of a text, and when we want to provide learners with feedback on their reading progress, it makes sense not to focus on their level of attention alone, but also to teach and encourage them to use reading strategies that optimize for knowledge gain. In this study, we explore the use of eye trackers to identify and track different reading patterns as suggested in the educational literature. We explore how quickly we can identify disadvantageous reading strategies, which gaze features help track reading strategies, and how such strategies relate to knowledge gain.

The next phase in developing the reading support system is to give students feedback on their reading progress. Over the past 40 years, intelligent tutoring systems have gained a lot of attention for supporting the learning process, as there has been an increase in the personalized and individual learning support that systems have been able to provide to users. The latest developments in intelligent tutoring systems have clearly proved that users of tutoring systems can make swift improvements and radically enhance their performance in precise areas and skills [2]. However, most of the support provided by intelligent tutoring systems has overwhelmingly been designed to support the learning of tasks that have a clear answer, such as solving a mathematical problem [31, 43, 72]. There are fewer systems that support open-ended problem solving tasks, such as essay writing or reading [3, 16, 78]. While most personalized-feedback systems are GUI-based, an increasing body of work also explores the use of social robots in educational settings [7, 47]. Compared to pure GUI systems, social robots have an embodiment. Studies show physical embodiment might be advantageous for learning. For instance, the use of a physically embodied over its virtual counterpart resulted in more compliance with its requests [5]. Additionally, robotic embodiment has been found advantageous for learning when solving cognitive puzzles [49] and losing weight [44]. It has been shown that the embodiment of a robot can impact student motivation positively [10]. All in all, social robots have been shown to be effective at increasing cognitive and affective outcomes and have achieved outcomes similar to those of human tutoring on restricted tasks [7] and the robotic embodiment has been linked to increased task performance [50] over their virtual counterparts.

Our long-term goal is to create an intelligent robot tutoring system that can evaluate learners in real-time and provide feedback when and if necessary. However, further research is still needed to determine whether physical robot intervention has a positive effect on participants' performance as well as the relationship between eye gaze patterns and knowledge gain levels.

To our knowledge, no previous study has yet explored whether using a robot in a reading task increases students' knowledge gain compared to a traditional GUI system. Therefore, we conducted research to investigate and compare participants' knowledge gain levels and gaze patterns in response to feedback from a GUI interface and a humanoid social robot.

Hence, in this paper, we pose the following research questions:

- How do knowledge gain and robotic feedback type relate to reading strategies? (R1)
- How quickly can we predict a student's knowledge gain based on gaze features and how do different models impact this prediction? (R2)

In the current paper, we contribute to the literature on the relationship between reading strategies and knowledge gain through the use of fixations, saccades, area of interest, and gaze coverage related features. Additionally, we contribute to learning on e-reading tasks by furthering our understanding of how feedback from humanoid social robots affects participants' reading strategies. In contrast to the previous study, which mainly focused on investigating children's learning patterns and the impacts of robots on them, we emphasized the college-aged adult's gaze pattern analysis when reading a full-text article. Finally, we contribute to exploring whether we could use reading strategies' related features to make predictions on the real-time simulation reading process from the beginning.

2 Background and related work

2.1 Analysis of reading competence in the educational sciences

Students' reading strategies are a reflection of their level of understanding or knowledge gain [27, 58]. The reading strategies show how readers conceive a task, what textual cues they pay attention to, how they interpret what they read, and what they do if they don't comprehend something [12]. A vast amount of research that concentrated on describing reading strategies involved in understanding has found that good readers are better at monitoring their comprehension than poor readers, that they are more aware of the strategies they use than poor readers, and that they use strategies more flexibly and efficiently [12, 22, 25, 63]. In this case, we may expect to see more of a range of reading patterns for readers with high learning gains compared to those with lower learning gains.

"Repair strategy" is one strategy that includes behaviors such as rereading and reading ahead [59]. "Extensive" and "intensive" reading are two other kinds of reading strategies, "extensive" reading consists of reading activities in which

the reader does the majority of skimming and scanning [29]. Skimming and scanning are reading strategies that make use of quick eye movements and keywords to skim through text quickly for slightly different objectives. Skimming is reading quickly to acquire a broad overview of the content. Reading quickly to locate certain data is known as scanning. Extensive reading can help readers find factual bits of information in the text but may not support them in developing a deeper understanding of the text. "Intensive" reading is reading in which the reader reads a page to identify the general meaning [62]. Intensive reading is useful to develop reading comprehension [61, 80], which is the process of creating meaning from text [23, 74]. Both extensive and intensive reading are needed for learners to fully benefit [28].

Another important reading strategy is the scanning technique, which looks for details relevant to questions that might have answers at the end of the assignment [4]. With the proper scanning methods, students' knowledge gain can be increased [1]. One of the most well-known techniques, made popular by the NNGroup eye tracking study [60], is the F-shaped scanning pattern. The F-shape pattern is a common scanning behavior exhibited by users when reading blocks of content, where they tend to read in a horizontal movement across the top, followed by a vertical movement down the left side of the page, forming the shape of the letter "F". Additionally, these strategies are something that can be learned [55].

In this paper, as the first step towards an interactive system, we are translating those reading strategies involving intensive and extensive reading, rereading, scanning techniques, and focusing on important sections of the text into gaze features and investigating their effect on knowledge gain.

2.2 Reading strategies related gaze features

The advancement of eye tracking technology has enabled us to recognize reading habits and strategies, as well as to assess cognitive load [70], by analyzing various features of the gaze. The most common signs of eye movement are those linked to fixations (points where the eye remains fixated at the same point) and saccades (eye movements between fixations) [82]. Many studies on eye tracking and reading tasks have used fixation and saccade related features to measure the correlation between students' eye movement and reading performance [36, 39]. "Intensive" reading can be indicated by longer fixation times [38, 39, 67]. The saccadic amplitude has been shown to be effective in distinguishing between reading and scanning texts [30]. Saccadic angles can be used to identify some scanning patterns, such as the F-shaped pattern [71], and can be used to distinguish reading from skimming [9]. Some studies categorize fixation into first-pass progressive fixation and regressive fixation [33, 37, 40]. Kaakinen and Hyona [35] suggested deeper reading comprehensive associated with longer first-pass fixation duration. "Repair strategy" can be indicated by the number and duration of regressive fixation [40].

Area of interests (AOIs) and gaze coverage are also employed by researchers to analyze reading patterns, which can be a good indicator of reading strategies. Fixations within the AOIs reveal the ability of the students to locate important information [14, 64, 83], which helps to evaluate the reading strategies. Gaze coverage shows a student's reading area size in a specified time window. Higher gaze coverage means more "extensive" reading is involved. [15] discovered that experts' gaze coverage during fashion designing was significantly higher than apprentices', implying that gaze coverage is another indicator to measure different reading strategies.

In addition, "total pass" or "regression path" reading time is a measure used to investigate processing difficulty [54]. In this study, we are also carrying out an array of explorative experiments to investigate which gaze features are proving to be helpful for the prediction of knowledge gain.

2.3 Robot tutoring

A growing body of research supports the potential of social robots as tutors or peer learners in ITS [75], and robots as learning companions are gaining more and more attention. From the perspective of robot embodiment, a variety of robots were used in ITS, from tiny toy-like robots to large android robots, but the Nao robot and Keepon robot were the most frequently used robots [7]. Social robots can interact socially because they all have humanoid features such as a head, eyes, mouth, arms, or legs.

The most common robot tutoring applications focus on learning computer science, physics, mathematics, and language [53]. In language learning courses, robots were frequently seen as human substitutes who could offer tutoring on a one-to-one basis at various paces [42, 51], or as social support who could keep students company in different settings [73]. In the reading use-case scenario, the robot acts as a reading companion in most studies. Results in many studies positively supported that the robotic reading companion enhanced participants' comprehension and made reading activities more interesting [6, 17, 26, 52, 81]. Yadollahi et al. [79] built a CoReader activity platform that uses a Nao robot as a reading companion for children aged 6–7, It was noteworthy that the robot's engagement could either be distracting or beneficial towards different children. Lin et al. [51] developed a service robot that can provide motivation, knowledgeable guidance, and a reading companion to help child patrons in library settings. However, the majority of studies focused on children, with adults receiving less attention. Kontogiorgos et al. [45] employed a Furhat robot as a conversational agent to investigate the impact of robot embodiment and social behavior on adults' grounding behav-

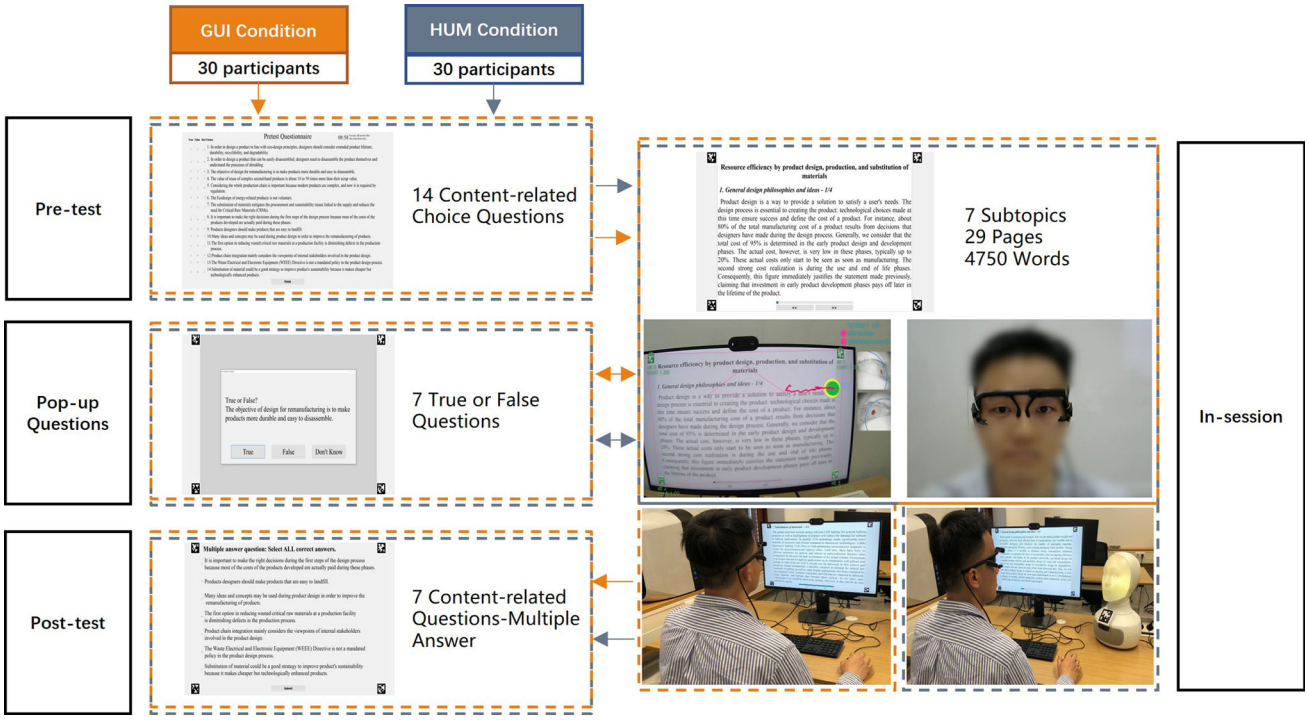


Fig. 1 Overview of experiment setting

iors during task-oriented dialogue. Bickmore et al. [8] created virtual agents to help researchers enhance their scientific presentations. Ince [34] developed a humanoid robot to play drums with adults, and the results revealed that the physical robot tutor, feedback, and training mechanisms improved the subjects' performance. According to what we understand, research for an adult reading companion robot is currently lacking. Therefore, in this paper, we focused on discovering the impacts of robot tutoring on college students during an English reading task.

3 The reading task case-study

We designed an e-reading experiment in which 60 volunteers were recruited to complete an English reading test in one of two settings (see Fig. 1).

3.1 Reading task

Participants engaged in an English reading comprehension task on "Waste management and critical raw materials", which was taken from a video course that explains how to increase resource efficiency through product design, manufacturing, and material replacement. The aim of choosing this text, in particular, was to make most participants unfamiliar with the topic, and the pre-test findings supported this. The average number of correct answers in the pre-test across all conditions was 2.97 ($SD = 2.97$, $Mdn = 2.37$) out of 14 questions. All the reading material is text, with a word count of 4750 evenly distributed over 29 pages, seven subtopics of similar lengths, and a reading period of 30 min. For the full text, visit https://github.com/hit-lx/JMUI/tree/main.

3.2 Manipulation

3.2.1 GUI condition versus HUM condition

Participants were randomly assigned to one of two conditions: a GUI system (GUI) with only a screen or a humanoid system (HUM) with both a screen and a humanoid social robot named Furhat. Both conditions involved a screen-based interaction. The HUM condition also included empathetic speech-based feedback provided by the Furhat robot, which was activated seven times, 2 s after beginning to read the final page of each subtopic.

3.2.2 Questions design

We designed two questions for each subtopic based on the content, totaling 14 questions. In the pre-test, participants must answer all 14 questions with "True", "False" or "I don't know" and they are encouraged to select "I don't know" if they are unsure about the answer. As for the post-test, seven questions are in the form of pop-up questions that appear when finishing one subtopic. Participants could

answer "True", "False" or "I don't know" to each question. The seven remaining questions, one related to each section, were transformed into answer options for a multi-select question, which was presented to the reader after all of the reading was completed. The reason behind the design of two types of questions is that we want to use two levels of difficulty to better evaluate participants' understanding of the context rather than their memory. See "Appendix A" for the full questionnaire.

3.2.3 Feedback design

In this study, there are two types of feedback: screen-based feedback and Furhat-based feedback. Screen-based feedback gives participants correct answers and explanations after completing each pop-up question. Furhat-based feedback, which is a type of speech-based feedback, was utilized in the furhat-based interaction to prompt participants to reflect on their understanding of the content. Feedback that is given with empathy and reflection has been associated with the cognitive, affective, and behavioral growth of students in the learning process, leading to positive experiences and successful learning results [48]. During the reading session, the Furhat robot will pose questions and provide empathetic and reflective feedback based on the responses of the participants. Feedback may come in the form of words, gestures, expressions, or LED blinks. See Table 1 for the protocol of robot feedback.

3.2.4 Knowledge gain assessment

The pre-test is conducted to measure the baseline knowledge of participants. The post-test questionnaire is to acquire knowledge after the reading session. We calculated the level of knowledge gain by comparing the number of correct answers in the pre-test and post-test. We divided all 60 participants into two groups by means of a median split. The first group comprised the high knowledge gain group, and the second group comprised the low knowledge gain group. The decision to refrain from separately comparing low and high knowledge gain groups in the two conditions was driven by the limited sample size, notable differences in average scores, potential score overlap across conditions, and the aim to maintain balanced group sizes for reliable analysis.

3.3 Experimental settings

3.3.1 GUI design

For this study, we developed an e-reading application for the GUI and the HUM conditions. The GUI system will guide participants to complete the entire experiment after calibration. To provide precise word-by-word detection, a 27-inch display with a resolution of 2560 × 1440 was employed, and the text in the main body was rendered at a size of 47 pt.

3.3.2 Eye tracking device

We chose the Pupil Core eye tracking device,[1] a wearable eye tracking device, for gaze feature collection, as we want to record gaze data on both the humanoid robot and the screen. It has a 30 Hz frontal camera and two 120 Hz infrared cameras focused on capturing the gaze and eyes of learners. The eye tracker has implemented software and a GUI that are platform-independent and include state-of-the-art algorithms for real-time pupil detection and tracking, calibration, and accurate gaze estimation. Gaze data is collected for each eye and includes the x and y positions of the gaze on the image captured by the frontal camera. Gaze position on the specified region is also available by using Apriltags to define the detecting surface. These data can be converted into a series of fixations. Results of a performance evaluation show that Pupil can provide an average gaze estimation accuracy of .6 degree of visual angle (.08 degree precision) with a latency of the processing pipeline of only .045 s [41].

3.3.3 Calibration

To accurately calibrate the eye tracker device, we used the screen marker mode; during calibration, four markers would appear sequentially in the four corners of the screen; participants had to hold their heads steady and only move their eyes to fixate on the maker. After calibration, gaze accuracy reaches .60 degrees.

3.4 Procedure

The reading was conducted with an eye tracking device and pupil lab core set for an average duration of 50 min, comprising 30 min of reading and 20 min of questionnaire and calibration. The experiment begins with instructions and a pre-test; participants will have 10 min to answer fourteen questions, though they typically use much less time. The calibration session will follow, during which we will assist participants in wearing the eye-tracker and calibrating it appropriately. Then participants can start the reading session whenever they are ready. In the reading session, participants have to read seven subtopics and their related questions. At the end of each subtopic, pop-up questions that ask the same questions as the pre-questionnaires 1–7 will be triggered. Unlike the GUI-based condition, the Furhat-based condition has a humanoid robot implemented. Reflective and empathic robot feedback will be provided 2 s after the last page of each

[1] https://pupil-labs.com/.

Table 1 Robot feedback protocol

Robot question	Participant response	Robot feedback: speech	Gesture	LED
Did you understand the content that you just read about?	Yes	My internal thumbs up for you! Keep up the good work!	Big smile + Nod	White blink (1.5s)
	No	No problem! We can always review once more!	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None
Did you understand everything in the text?	Yes	That's music to my ear! Let's move on	Big smile + Nod	White blink (1.5s)
	No	I know, it's all about learning. We can go through the unclear part once more	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None
Did you like the text?	Yes	I am happy to hear that. I am interested in this topic, too	Big smile + Nod	White blink (1.5s)
	No	I am sorry to hear that. I believe that you will like the next subtopic better	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None
Have you been focused while you were reading?	Yes	That's super! Let's try to keep your good focus until the end of the text!	Big smile + Nod	White blink (1.5s)
	No	Maintaining good focus is always difficult. You are doing good already	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None
Are you clear about the subtopic that we just went through?	Yes	That's amazing! I am proud to be your reading companion	Big smile + Nod	White blink (1.5s)
	No	It's okay. I always review once more if I don't understand something	Big smile	White blink (1.5s)
		Maybe you can do it too		
	N/A	Alright, Let's continue	Smile	None
Do you think you can apply the knowledge that you just learned?	Yes	Wow, you are a fast learner!	Big smile + Nod	White blink (1.5s)
	No	I know it's not that easy. One tip is to reflect on the main point while you are reading	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None
Can you recall the main point of the subtopic in your mind?	Yes	Great, you are doing even better than I expected!	Big smile + Nod	White blink (1.5s)
	No	Recalling new information always takes. some time. It's all about practice	Big smile	White blink (1.5s)
	N/A	Alright, Let's continue	Smile	None

subtopic is triggered. Once the reading was finished, the post-test questionnaire, which was the same as the pre-question 8–14, was given.

3.5 Participants

In order to reach our target of 60 college students for the e-reading task, we actually enrolled 78 participants due to a malfunction of the eye tracking device. Participants were non-native speakers of English but students who use English on a daily basis for their educational activities. For technical reasons related to the workings of the wearable eye-tracking device, we only included students who did not wear glasses or contact lenses. We advertised the experiment through social media and through direct advertisements on campus, and only university students were eligible to participate in the experiment since we use the university's recruitment system. Participation in the experiment was voluntary, and participants were compensated with a €10 voucher. The GUI condition included 18 males and 12 females ranging in age from 19 to 33 ($M = 25.8$, $SD = 3.35$). The HUM condition included 19 males and 11 females ranging in age from 19 to 37 ($M = 24.1$, $SD = 4.30$).

3.6 Hypotheses

Based on the above-mentioned literature, features related to fixations, saccades, AOIs, and gaze coverage appear to be effective in indicating reading strategies, such as intensive or in-depth reading indicated by more and longer fixations, lower saccade amplitude, shorter saccade duration, higher gaze coverage, and the scanning strategies suggested by the saccade angle related features. The reflective nature of a humanoid social robot could be beneficial to reading comprehension. As a result, this study was built on the premise that features related to fixations, saccades, AOIs, and gaze coverage can effectively distinguish participants' knowledge gain levels and that participants exposed to social robots' feedback will change their reading patterns, affecting knowledge gain.

Therefore, we hypothesize that:

- **H1** Participants in high knowledge gain group will have reading patterns with more and longer fixations, lower saccade amplitude, shorter saccade duration, as well as a higher gaze coverage, indicating that they were reading intensively;
- **H2** Longer fixations on AOIs and the application of repair procedures will boost knowledge gain;
- **H3** Scanning strategies have an effect on knowledge gain.
- **H4** Empathetic feedback delivered by a social robot compared to a baseline condition will result in a more in-depth reading.

4 Feature engineering

The purpose of the extraction of features from gaze data was to uncover differences in reading patterns between groups and essential features to distinguish different levels of knowledge gain and the impact of robot feedback on participants' reading behaviors and knowledge gain levels. Based on the literature review and previous analysis, we chose the features listed in Table 2 to serve this purpose.

4.1 Data pre-processing

During data collection, the border of the text-showing region was marked with four AprilTags [57] to define it as a surface. We only used gaze data located on the surface for feature extraction. The raw gaze data contains a lot of spikes and outliers. We first apply the Savitzky-Golay filter [69] to the gaze point coordinate data for smoothing and use the 1.5 interquartile range rule to eliminate outliers. We also attempted to use standard deviations as a criterion to detect and eliminate outliers. However, our data had a very small proportion of extreme outliers with values greater than 100 times the mean, making this strategy ineffective in our case.

4.2 Primary-level gaze feature extraction

According to the previous literature [11, 19, 20, 32], fixation and saccade associated features play a crucial role in distinguishing different reading strategies.

4.2.1 Fixations

Fixations were extracted by the pupil lab application under the definition of fixations being consecutive gaze points within a range of 1.5 deg of visual angle for between 100 and 800 ms, 100 ms is the shortest duration for naturalistic eye movements during reading [30, 66]. 800 ms is the typical maximum duration for college-aged readers [65]. In this study, we employed a time window to collect features such as the sum and average of the number and length of fixations, as well as the dispersion of fixations.

4.2.2 Saccade features

We compute saccades from fixations based on the definition that saccades are quick, simultaneous movements of both eyes between two or more phases of fixation in the same direction [30]. We excluded saccades that lasted less than 100 ms, the shortest fixation duration, to eliminate the measurement error caused by fixation detection. Regarding saccade features, we compute saccade duration, saccade amplitude, saccade velocity, and relative and absolute saccade angle distributions. The saccade duration was measured as the num-

Table 2 Summary of features

Feature	Standard or common unit	Description
Gaze coverage (GC)	Sentence	The proportion of the image which is gazed at by the individual at least once
Fixation duration (FD)	Word	Elapsed time in ms of fixation
Saccade duration (SD)	Region	Elapsed time in ms of saccade
Saccade amplitude (SA)	Region	Distance of saccade in pixels
Saccade angle absolute (SAA)	Region	Angle in degrees between the x-axis and the saccade
Saccade angle relative (SAR)	Region	Angle of the saccade relative to previous gaze point
Saccade velocity (SV)	Region	Saccade length/saccade duration
Fixation number (FN)	Region	Total of number of fixation in one page or a time window
Fixation dispersion (FDP)	Page	Root mean square of the distances of each fixation to the average fixation position
Horizontal saccade proportion (HS)	Page	The proportion of saccades with relative angles ≤ 30 degrees Above or below the horizontal axis
Fixation saccade ratio (FSR)	Page	Ratio of fixation duration to saccade duration
Fixation duration (FD)	page	the total duration of Fixations in one page
Progressive fixation duration (PFD)	Page	The total duration of first-pass fixations in one page
Regressive fixation duration (RFD)	Page	The total duration of second-pass fixations in one page
Regressive fixation number (RFT)	Page	The total regressive fixation number in one page

Region: word, phrase, clause, sentence, page. Bold indicates the mean, median, min, max, std. dev., range, kurtosis, and skew of the distribution of each measurement were used as features

ber of seconds between two consecutive fixations, whereas the saccade amplitude was the length of the saccade, which equals the distance between the fixations. In this study, we used the number of pixels on the saccade trajectory between two subsequent fixations to compute the saccade amplitude, given that the fixation coordinates were standardized so that the maximum amplitude value was below 1. The saccade velocity was calculated as the saccade amplitude divided by the saccade duration. The absolute saccade angle is defined as the angle between the line segment between two subsequent fixations and the x-axis. The relative saccade angle was computed as the angle between two subsequent saccades.

4.3 Secondary-level gaze feature extraction

4.3.1 Gaze coverage

Gaze coverage is a metric used in the area of eye tracking but has been defined in different ways [24, 30]. In our case, the gaze coverage was defined as the percentage of the text that the participants looked at at least once [56]. We divided the text region into a regular grid of size 20 × 10 based on the size of the GUI region of text. This set of grids is defined by the layout of the text area. Most of the cells in the grid contain one or a few short words, as shown in Fig. 2a. We counted the number of gaze hits, frequency of gaze hits with cells [68], towards each cell of the grid for each participant in a given time window.

4.3.2 Secondary fixation features

We conducted a phrase-level analysis of the gaze patterns of the participants based on the grid that we built to calculate the gaze coverage. We divided the fixations into two categories: first-pass and second-pass fixations. First-pass fixations are those made when reading through the target cell of the grid for the first time, whereas second-pass fixations are directed back to the target cell of the grid from a subsequent phrase (i.e., after its initial processing is completed) [40]. Based on the aforementioned definition, we compute the progressive fixation time by summing up the forward-going fixations that landed on an unread part of the phrase. The look-back fixation time was calculated by adding the duration of the fixations made after the first-pass reading. The progressive fixation time and look-back fixation time were calculated from gaze data on each page of text.

4.3.3 AOIs coverage

In our case, AOIs were manually designated as areas containing important passages (that is, sentences or phrases containing information that leads to correct responses to questions), as shown in Fig. 2b. To work with fixation duration, it is necessary to connect the eye tracker data with the text region. We manually specified the parts of the GUI text image and got the image pixel coordination of those regions. Afterward, we label every fixation depending on whether it hits those areas or not. Hence, we obtain the fixations that are located in those areas. We extracted fixation features based on gaze data for each page. We categorized every fixation into four groups based on whether it is in the AOIs and whether it is a progressive fixation.

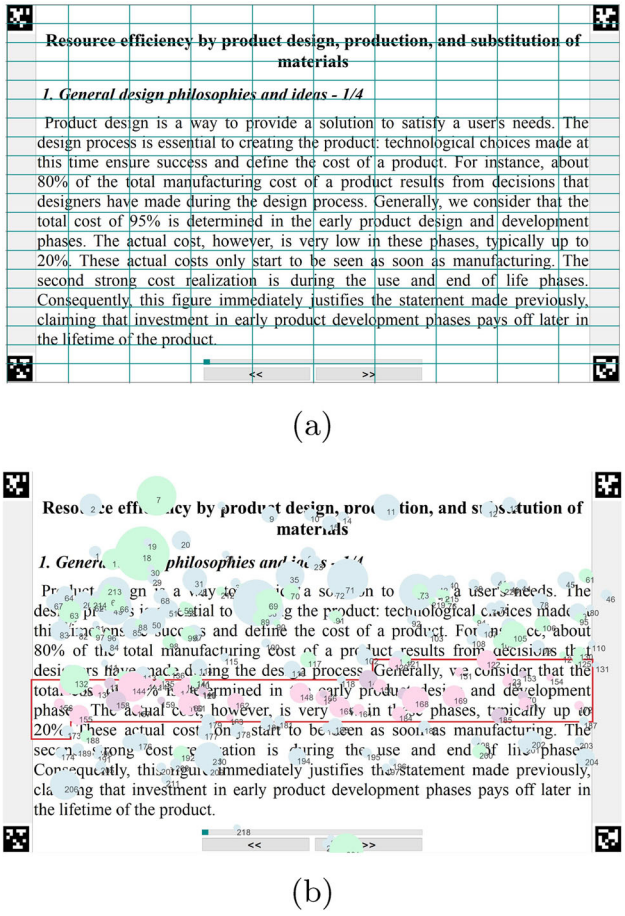


Fig. 2 Visualization of reading surface annotation: **a** shows the 20 × 10 gaze grid on the reading surface. **b** lists four types of fixation distribution. size of spot represents fixation duration; AOIs are circled by boxes; colors for fixation category, size of spot represent fixation duration, colors for fixation category (light blue: fixation is progressive fixation but not in the AOIs, light green: fixation is regressive fixation but not in the AOIs, purple: fixation is progressive fixation and also in the AOIs, pink: fixation is regressive fixation in the AOIs)

4.3.4 Miscellaneous gaze properties

According to [19], the miscellaneous gaze properties that are related to saccades and fixations were also deployed and could be promising features to differentiate two knowledge gain level groups' reading strategies. These consisted of the

horizontal saccade proportion, fixation dispersion, fixation duration/saccade duration ratio, and the number of fixations. The horizontal saccade proportion was the percentage of saccades that were less than 30 degrees above or below the x-axis, indicating the participants' reading pattern from right to left. The root mean square of the distance between each fixation and the average fixation on each page was used to calculate fixation dispersion, which was utilized to reveal the spatial distribution of fixations on each page. The fixation duration to saccade duration ratio was calculated as the ratio of the total duration of all fixations to the total duration of all saccades, indicating the percentage of intensive and extensive reading.

5 Feature analysis results

A series of Multivariate Analysis of Variance (MANOVA) [76] experiments were conducted to analyze differences between different groups and conditions in the hypothesis-related features. For the post-hoc test, we used Tukey's HSD test.

$H1$: To test our hypothesis on the high knowledge gain group's gaze pattern, a one-way multivariate analysis of variance (MANOVA) was conducted to compare the two groups in terms of their fixation number and duration, saccade amplitude and duration, and gaze coverage. A statistically significant MANOVA effect was obtained ($Pillai'sTrace = .016$, $F(5, 1732) = 5.44$, $p < .001$), which indicates that knowledge gain level has a statistically significant association with those five features. The post hoc test revealed significant differences between all those five features (adjusted p-value $<.001$). Using Tukey's HSD Test for multiple comparisons, we found that the mean value of fixation number ($p < .001$, $95\%CI = [-35.30, -7.14]$), fixation duration ($p < .001$, $95\%CI = [-5.60, -1.40]$), saccade amplitude ($p = .003$, $95\%CI = [-.004, .017]$), saccade duration ($p < .001$, $95\%CI = [.004, .018]$), as well as gaze coverage ($p < .001$, $95\%CI = [.03, .047]$) was significantly different between the high knowledge gain group and the low knowledge gain group. The results indicate the high knowledge gain group had fewer fixations, shorter average fixations, greater saccade amplitude, longer saccade duration, as well as lower gaze coverage compared to the low knowledge gain group. The results revealed that while the number and duration of fixations, as well as saccade amplitude and duration, contradicted our hypothesis, the difference in gaze coverage was matched.

$H2$: We analysed the sum of progressive fixation duration ($APFD\_sum$), average of progressive fixation duration ($APFD\_avg$), progressive fixation number ($APFN$), sum of regressive fixation duration ($ARFD\_sum$), average of regressive fixation duration ($ARFD\_avg$), regressive fixation ($APFN$) number in AOI of two knowledge gain level group. A Multivariate Analysis of Variance (MANOVA) was conducted to identify a possible difference between the groups. The Pillai's Trace test statistics were not statistically significant ($Pillai'sTrace = .071$, $F(6, 53) = .677$, $p = .668$) and indicate that knowledge gain level has no statistically significant association with the combination of those features in AOIs. The post-hoc test revealed that none of the feature's adjusted p-values are below. 01. Thus, the hypothesis (H2) is false.

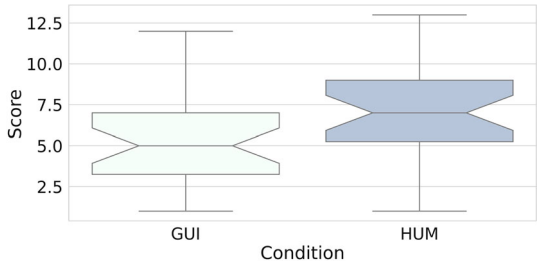


Fig. 3 Knowledge-gain between GUI and HUM

$H3$: We used the same method to analyze the difference in saccade angle related features, which are saccade absolute angle, saccade relative angel, and horizontal saccade proportion, between the two knowledge gain groups. The MANOVA results ($Pillai'sTrace = .053$, $F(5, 1727) = 19.324$, $p < .001$) revealed that there is a statistically significant association between knowledge gain and the saccade angle related features, which indicates scanning strategies. The following post-hoc test revealed significant differences between average saccade absolute angle ($p < .001$, $95\%CI = [-.066, -.012]$), and horizontal saccade proportion ($p < .001, 95\%CI = [.023, .046]$). The results show that the high knowledge gain group had more horizontal saccades and a larger absolute saccade angle, which can be interpreted as a line-to-line reading pattern, often known as an F-shaped scanning pattern, that aids in knowledge gain. Therefore, H3 is true.

$H4$: There are 30 participants for each condition, and 14 questions to evaluate their knowledge gain. The average number of correct answers across all conditions was 6.18 ($SD = 2.97$, $Mdn = 6$) out of 14 questions, with the GUI condition receiving 5.53 ($SD = 2.81$, $Mdn = 5$) and the HUM condition receiving 6.83 ($SD = 2.99$, $Mdn = 5$). As demonstrated in Fig. 3, the participants performed better in the HUM condition than in the GUI condition, indicating that feedback from a humanoid social robot had a positive impact on the participants' reading comprehension.

The effect of knowledge gain levels and conditions on reading depth-related features such as fixation number, fixation duration, gaze coverage, and regressive fixation was investigated using a two-way MANOVA analysis. The results showed a statistically significant interaction

effect between the condition and the knowledge gain group ($Pillai'sTrace = .027$, $F(7, 1730) = 6.733$, $p < .001$).

Planned post-hoc comparisons of the interaction indicated fixation number ($p < .001$, $95\% CI = [16.602, 44.384]$), average fixation duration ($p < .001$, $95\% CI = [5.755, 9.833]$), gaze coverage ($p < .001$, $95\% CI = [.030, .047]$), regressive fixation duration ($p < .001$, $95\% CI = [2.115, 5.378]$), regressive fixation number ($p = .003$, $95\% CI = [22.549, 342.663]$) had a significant difference between the conditions at post-test. However, results in the saccade amplitude and saccade duration showed no difference between the GUI condition and the HUM condition. The results revealed that the HUM condition had higher average gaze coverage, indicating that the reader in the HUM condition covered more content in the same amount of time. One explanation is that the participants in the HUM condition skimmed and skipped more in favor of finding the main point. In addition, the HUM condition had a longer average fixation time and a higher rate of regressive fixation, meaning that each fixation lingered longer than the GUI condition and that more "repair" strategies were used.

To further understand the immediate impact of social robot feedback, we conducted a one-way MANOVA to compare the regressive and progressive fixation related features along with average gaze coverage extracted from the reading of the last page of each subtopic, which was the page that social robots gave feedback on, with other pages. The result revealed that there was a statistically significant difference between the two groups of features ($Pillai'strace = .029$, $F(7, 802) = 3.363$, $p = .002$). The results of the post hoc test only showed significant differences in regressive fixation duration ($p = .006$, $95\% CI = [.0, .009]$), which indicates that under the HUM condition, normal reading (reading without feedback) had a shorter average duration of the regressive fixation. Those findings indicated a spike in fixations, particularly regressive fixations, on the last page of each topic when feedback was given, implying that the feedback prompted readers to reread the text whenever their comprehension was lacking. This result rules out the possibility that the social humanoid robots' presence made participants adopt more "repair" strategies.

6 Supervised classification

For the purpose of establishing the reading support system, we aim to build supervised machine learning models that can distinguish knowledge gain levels and achieve real-time knowledge gain prediction, the entire process depicted in Fig. 4. We investigated and tested a variety of classifiers, including logistic regression, ridge, support vector machine, bagging, extra tree, Gaussian process, AdaBoost, GaussianNB, SGD, passive aggressive, k-nearest neighbors, decision tree, random forest, fully connected neural network, linear discriminant analysis, and random tree.

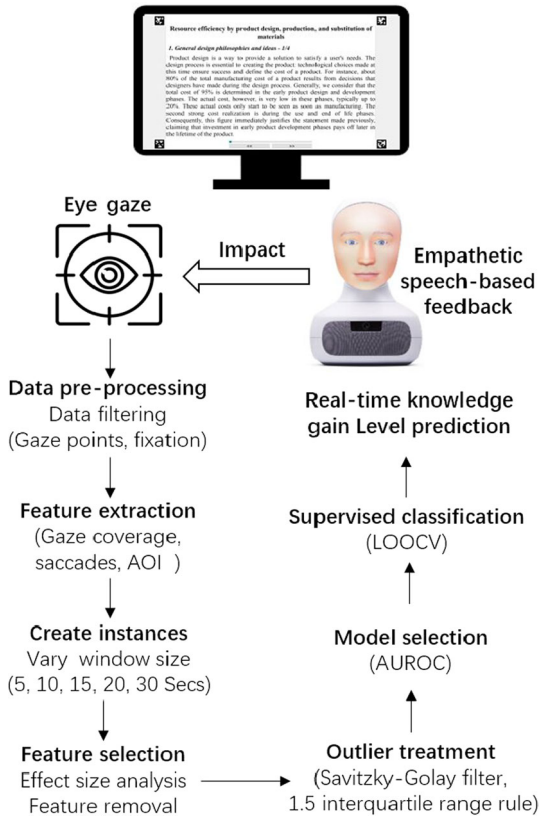


Fig. 4 Visualization of the machine-learning approach outlined in the supervised classification section

6.1 Feature extraction and selection

The descriptive features of the eye movement were gaze coverage, fixation duration, saccade duration, saccade amplitude, saccade velocity, and relative and absolute saccade angle distributions. With the goal of finding the best feature extraction window for the classifiers, we computed fixation duration, saccade duration, saccade amplitude, saccade velocity, and relative and absolute saccade angle distributions with a set of time windows of 5, 10, 15, 20, and 30 s, as well as a window of one page of the reading text. For each of these eye movement measurements, we computed the minimum, maximum, mean, median, standard deviation, skew, kurtosis, and range.

We did not include some features that are strongly related to other features, such as the number of saccades, since saccades are the rapid eye movements between fixations. We extracted 15 types of primary-level and secondary-level features and obtained 119 gaze features for further analysis and feature selection.

Table 3 Means (with standard deviations in parentheses) and effect sizes for gaze features corresponding to two knowledge gain group

Feature	Low knowledge gain	High knowledge gain	Cohen's d	*P* value
HS	.225 (.061)	.255 (.086)	− .393	< .001
FN_W_max	44.203 (8.239)	41.536 (7.652)	.337	< .001
SD_skewness	1.129 (.241)	1.046 (.28)	.314	< .001
RFD	17.719 (12.321)	14.472 (9.542)	.299	< .001
FD_std	48.257 (12.279)	44.578 (12.38)	.298	< .001
SAA_min	− 2.803 (.536)	− 2.631 (.676)	− .278	< .001
FD_range	197.173 (52.555)	182.651 (53.934)	.272	< .001
FD_max	297.205 (52.548)	282.687 (53.917)	.272	< .001
SAA_range	5.512 (1.145)	5.169 (1.363)	.270	< .001
FD_W_median	165.145 (18.699)	160.066 (18.895)	.270	< .001
SAA_W_median	− .095 (.224)	− .157 (.241)	.267	< .001
FN_W_avg	37.687 (5.906)	35.911 (7.318)	.264	< .001
FD_W_min	147.076 (14.452)	143.216 (15.174)	.260	< .001
FD_W_avg	165.856 (18.819)	160.965 (19.292)	.256	< .001
SAA_max	2.71 (.631)	2.538 (.72)	.252	< .001
FN_W_range	12.956 (8.221)	11.266 (5.354)	.249	< .001
FN_W_median	37.598 (6.437)	35.906 (7.364)	.243	< .001
FN	249.743 (114.778)	224.457 (95.184)	.242	< .001
SD_kurt	.633 (.687)	.47 (.69)	.237	< .001
FD_avg	157.177 (16.406)	153.196 (17.247)	.236	< .001
SD_median	.068 (.027)	.078 (.055)	− .235	< .001
GC_max	.381 (.108)	.357 (.097)	.234	< .001
SD_W_range	.272 (.239)	.401 (.741)	− .225	.001
SAA_std	1.315 (.386)	1.221 (.446)	.224	.001
SA_W_range	.366 (.239)	.471 (.606)	− .221	.001
SAA_W_avg	− .102 (.199)	− .15 (.237)	.218	.001
SAR_kurt	− .972 (.337)	− .879 (.489)	− .218	.001
SA_W_std	.099 (.067)	.13 (.184)	− .213	.002
SD_W_max	.328 (.264)	.456 (.772)	− .212	.002
SA_W_max	.399 (.252)	.505 (.634)	− .212	.002
SD_avg	.084 (.036)	.097 (.071)	− .210	.002
SA_W_avg	.159 (.107)	.198 (.242)	− .206	.003
FN_W_std	3.41 (2.788)	2.955 (1.63)	.205	.004
SD_W_std	.073 (.075)	.11 (.229)	− .204	.004
PFD	24.339 (10.46)	22.413 (8.686)	.202	.004

HS horizontal saccade proportion, *FN* fixation number, *SD* saccade duration, *RFD* regressive fixation duration, *FD* total fixation duration, *SAA* saccade absolute angle, *SAR* saccade relative angle, *PFD* progressive fixation duration, *SA* saccade amplitude, *_W* denotes that features were extracted from a time window

To investigate the correlation of features with knowledge gain, we computed the effect size (Cohen's d), as [19], and performed a t-test for each feature to generate P values, which were then corrected with a Bonferroni correction [77]. Table 3 lists all 37 features in descending order of effect size, with absolute values greater than .2 and P-values less than .05. To sum up, the 37 features represent horizontal saccade ratio, saccade duration, saccade relative angle, saccade absolute angle, fixation dispersion, gaze coverage, fixation duration and number, saccade amplitude, and progressive fixation duration.

6.2 Model selection

Model selection was based on the use of a repeated hold-out validation method to evaluate classification models. This method ensured that data from each participant was exclusive to either the training or test set. We randomized the sequence of participants into two conditions, respectively.

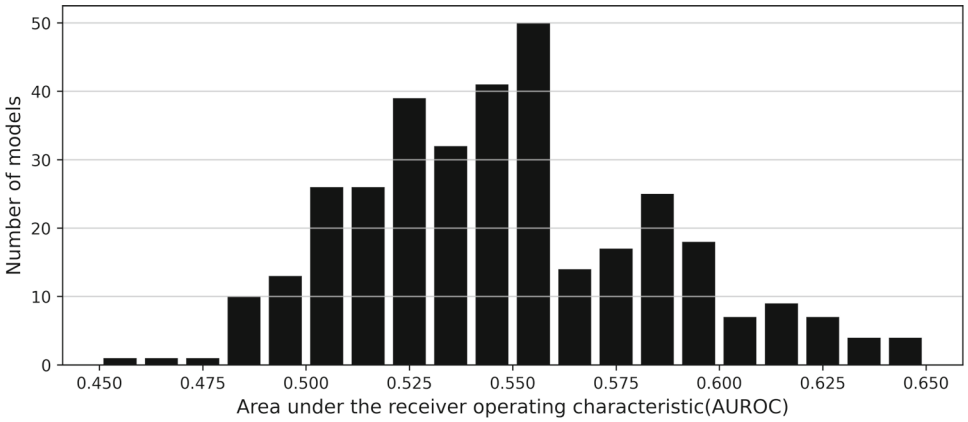


Fig. 5 Histograms of area under the receiver operating characteristic curves

For each fold, we chose 20% of participants' data from two conditions in sequence for the testing sets, while the data from the remaining participants was used to train the model. The process was repeated ten times to ameliorate the variance caused by random selection of participants. To resolve the class imbalance, the training set's class distribution was equalized by downsampling. The training set was subjected to feature selection in order to identify the most diagnostic features and avoid overfitting.

The histogram of the AUROCs (areas under the receiver operating characteristic [ROC] curve) for each of the 345 candidate models is shown in Fig. 5. A total of 92.17% of the models had ROCs greater than chance ($AUROC >$ .50). The 4 best models (each with an AUROC above .64) were linear discriminant analysis (LDA), k-nearest neighbors (KNN), Gaussian process classifiers (GPs), and support vector machine with RBF kernel (NuSVC). Notably, all four best models were with a window size of 20 s, and used a total of 35 features after feature selection. The overall best performance was achieved by a support vector machine with an AUROC of .649, which reflects a 29.8% improvement over a chance model ($AUROC = .50$).

To improve the performance of machine learning models, we use a soft voting classifier to make predictions based on the findings of the four best-performing models. The same repeated holdout validation method was used to compare the train and test accuracy of four models and the voting classifier. The procedure was repeated 50 times to get the average train and test accuracy (see Fig. 6). As the voting classifier achieved the highest accuracy (74.2%) and the second highest train accuracy (81.1%), we employed this voting classifier for the next experiment.

6.2.1 Feature ablation study

We also performed two feature ablation studies on the voting classifier to verify the inclusion of each type of gaze feature and certain features related to the reading strategy

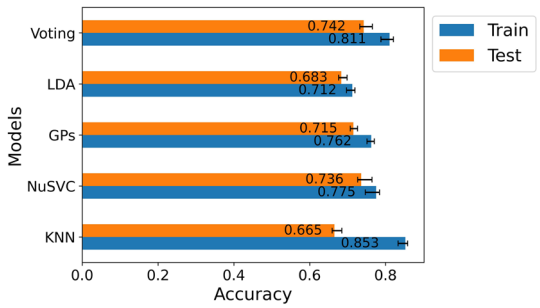
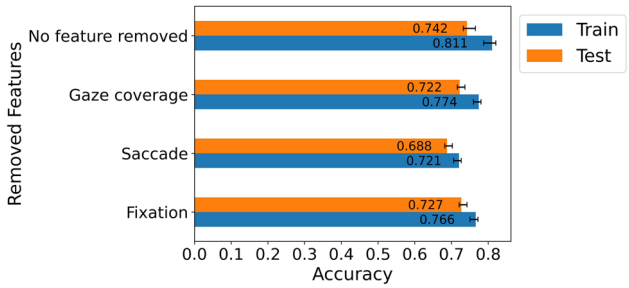
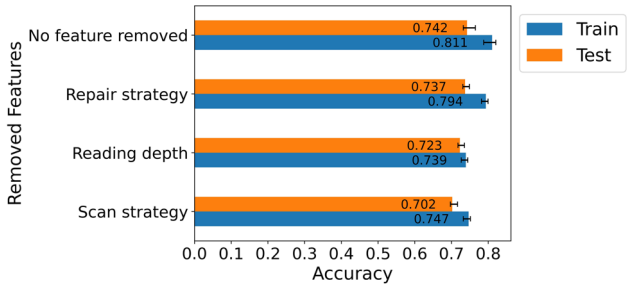


Fig. 6 A comparison of the performances of different classification models balanced accuracy averaged over instances trained on randomly selected 40 participants of data (95% CIs as error bars) (*Voting* votingclassifier, *LDA* linear discriminant analysis, *GPs* Gaussian process classifier, *NuSVC* support vector machine with RBF kernel, *KNN* k-nearest neighbors classifier)

in the model training set. We removed one set of features at a time and compare the accuracy trained and tested with the model trained on all the features (see Fig. 7). In the first study, we categorized the 35 features into fixation-related features, saccade-related features, and gaze coverage-related features. We excluded the AOIs-related features from the ablation study because there was no significant difference in knowledge gains between the two groups, and the 35 features did not include AOIs-related features. Removing one set of features one at a time showed a slight decrease in training and testing accuracy compared to the model trained on all the features. According to the Kruskal-Wallis H-test, there is a significant difference in training ($\chi^2(3) = 701.76$, $p < .001$) and testing accuracy ($\chi^2(3) = 5.65$, $p < .001$) between feature sets. The results of Dunn's Multiple Comparison post hoc test revealed a significant decline in testing and training accuracy when removing gaze coverage related features (post hoc Dunn's $p < .001$), saccade related features (post hoc Dunn's $p < .001$), and removing fixation related features only caused a significant decline in training accuracy (post hoc Dunn's $p < .001$). In the second study, we classify the 35 features according to their relevance to different reading strategies, like "repair strategy", indicated by regressive fix-



(a)



(b)

Fig. 7 Ablation study results: accuracy of models trained on different sets of features with 95% CIs over models trained on different random samples as error bars. "No features removed" below on the y-axis shows a baseline model with all features included. Other labels mean a feature being removed from the training sample

ation; "scanning strategy", implied by the saccade angel and horizontal saccade proportion; and extensive and intensive reading ratio, or reading depth, shown by fixation number, fixation duration, and saccade velocity. A decline in training and testing accuracy when removing one set of features was spotted again. A Kruskal-Wallis H-test showed a significant difference in training ($\chi^2(3) = 703.21$, $p < .001$) and testing accuracy ($\chi^2(3) = 11.85$, $p = .008$) between feature sets. The only significant decline in testing and training accuracy was when removing scan strategy related features (post hoc Dunn's $p = .003$). Train accuracy was also significantly lower when removing any feature set when compared with the model trained on all the features (post hoc Dunn's $p < .005$).

6.3 Real time knowledge gain prediction

To find out how quickly we can predict students' knowledge gain (R2), we used the voting classifier to make predictions with the gaze features fed in page by page sequentially (29 pages in total). The leave-one-participant-out validation method was employed to evaluate the model; this method ensured that data from each participant was exclusive to either the training or test set. We use 59 participants' features as the training data and one participant's feature as the test set. The process was repeated 60 times until all 60

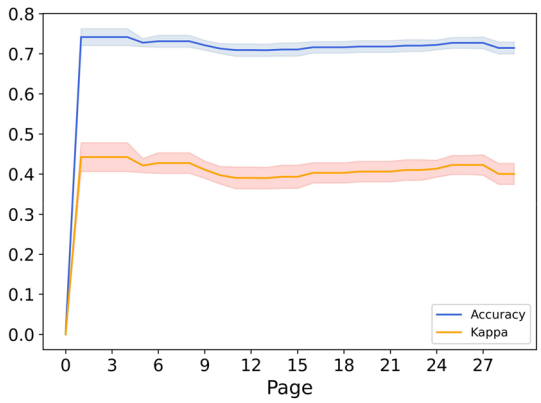


Fig. 8 Accuracy and Kappa value curve in real time knowledge gain group prediction (95% CIs as error band)

participants had been in the test set once. Data fusion and decision fusion were used to achieve the desired results. The classifier made knowledge gain predictions based on the participants' reading patterns on every subtopic. The features of one-page reading and consecutive several-page reading were fed into the voting classifier. We made the final prediction of the knowledge gain on this subtopic based on the "majority vote" decision fusion strategy. The overall real-time knowledge gain was decided by the prediction results of all the subtopics that have fed in. As data is fed in, Fig. 8 displays accuracy and a kappa value curve. The accuracy begins at 74.1%, fluctuates slightly in the middle, and ends at 71.5%. The kappa value, which began at .442 and finished at .400, exhibits a similar tendency. This finding suggests that we can predict the knowledge gains of participants based on how they read the first page.

7 Discussion

Are There Significant Differences in Terms of Reading Pattern Between the Knowledge Gain Groups?

We need to reject H1. According to the data analysis, the high knowledge gain group had significantly fewer and shorter fixations, larger saccade amplitude, and more horizontal saccade proportion, contradicting our hypothesis. Fixations were fewer and shorter in the high knowledge growth group, implying that there is no direct correlation between fixations and knowledge gain. Furthermore, the scanning techniques of the high knowledge gain group, indicated by saccade features, demonstrated a positive relationship between the F-shaped scanning pattern and the performance of the reading task, proving our H3. When the analysis results are combined, it demonstrates that a better reading pattern for the reading task was to use horizontal scanning and skimming more frequently and only fixate on the areas that were considered relevant. This pattern is

interpreted as detailed scanning and skimming to locate all important parts of the text, with just intense reading on the important parts, allowing participants to read more efficiently and concentrate on the reading task. The findings revealed that scanning and skimming from line to line while focusing solely on the major parts of the text improves knowledge.

The reading pattern analysis in the AOIs shows no statistically significant differences between the two knowledge gain groups in the duration and number of progressive and regressive fixations, which rejects H2. This result indicates that, despite reading the questions during the pre-questionnaire session, all participants may not have remembered them. As a result, both groups failed to notice the paragraph that contains the answers to the questions or put extra effort into this part of the text. Although both groups exhibit the same fixation patterns on the AOIs, we believe the high knowledge gain group's more effective reading strategies allow them to locate all potentially important paragraphs in the text. However, due to the length of the text (30 min of reading), they were unable to focus their attention on all potentially important paragraphs.

In conclusion, significant differences exist across scanning strategies and extensive and intensive reading distributions among two groups (R1). The high knowledge gain group uses an F-shaped scanning method, focusing solely on the most relevant areas, which implies they read more extensively rather than intensively. The low knowledge gain group, on the other hand, read more extensively across the text and scanned between lines and sections more frequently. This is in line with [28] who revealed that a balanced approach of extensive and intensive reading will provide learners with maximum benefit.

What was the impact of the feedback from Furhat robot?

We expected longer fixations, lower saccade amplitude, lower gaze coverage, and better knowledge gain in the HUM condition. Our results show that participants in the HUM condition had wider gaze coverage, more and longer fixations in general, and more and longer regressive fixations in particular. Moreover,in the HUM condition, a surge in fixations was recorded, particularly regressive fixations on the last page of each topic when feedback was delivered. The higher gaze coverage indicates that people in HUM condition tend to use extensive reading to acquire the general idea of the text first. We also found that participants in the HUM condition reread the parts they found important. This is indicated by the increase in the number and duration of regressive fixations. Furthermore, the participants in the HUM condition performed better in finding the correct answers to the knowledge gain questions and fixated longer than participants in the GUI condition in general. Those findings showed that the Furhat robot's feedback effectively reminded participants to focus on their reading goal, find the answers to the questions in the pre-questionnaire, and encourage them to use the "repair strategy". To summarize, the robot's feedback significantly impacted the entire reading process. Participants were more task-focused and followed the instructions to reread the text whenever they found their understanding unsatisfactory (R1).

How do our results compare to previous research on the relationship between fixations, saccades, and high/low knowledge gain?

We compared our results in H1 to those of [11, 19], which conducted an experiment in a comparable setting. All three experiments involve college students finishing a computerized reading of a similar-length article. However, their study concentrated on the detection of mind-wandering, and the level of text comprehension was used as a baseline for the evaluation of mind-wandering. Their research showed that mind-wandering negatively related with comprehension scores, and the normal reading without mind-wandering has more fixation and larger saccade amplitude, but short fixation duration and saccade duration.

Our findings suggest that those in the high knowledge gain group had fewer fixations and shorter fixation durations but stronger and longer-lasting saccades. The main discrepancy in the results was related to fixation number and saccade duration. We examined the experimental setups of the three studies, as there may have been some distinctions that could have impacted the outcomes. The first distinction is that our study used the same material to form the questions in the pre- and post-tests, with the topic-related pre-test informing participants to locate the content in the text that provided the right answers to the questions that would follow. This difference could make the reading session a target-clear task, with those who are adept at skimming and recognizing the important details performing better on the pop-up questions and post-test questions. The second difference is that participants in this study were asked to report any mind wandering that occurred during the reading session. Both self-report and thought probes would disrupt reading consistency and cause rapid eye gaze shifts, which will increase fixation and amplify saccade amplitude.

How quickly can we predict a student's knowledge gain based on gaze features and how do different models impact this prediction? (R2)

We investigated whether the level of knowledge gain could be predicted using our reading strategy-related features. The models that performed the best in terms of test accuracy were the soft voting classifier, which was the ensemble of NuSVC, linear discriminant analysis, k-nearest neighbors, and Gaussian process classifier, with an average accuracy of 74.2% after 50 repeated holdout validation tests. This result not only shows that the classification model was able to predict a participant's knowledge gain based on the gaze features of one page, but it also demonstrates the strong predictive power of the selected features. Based on the feature

selection results, we used features associated with fixations, saccades, and gaze coverage to feed the models but excluded AOIs-related features. The feature ablation studies showed significant changes in model performance when removing scan strategy related features (saccade angle), gaze coverage related features, and saccade related features. These results support the hypothesis that the saccade-related features were most distinguishable, and using scanning strategies to predict knowledge gain was the most effective way. Removal of other set of features from the training set of the model did not show significant changes to the original model. It was unexpected that no significant changes occurred when the fixation-related features were excluded, despite the average training and testing accuracy declining. This can be explained by the fact that the fixation-related features are correlated with the selected features, which gives the classifier enough information about the fixation features. Moreover, most of the reading depth and "repair strategy" related features were related to fixation. This could be the reason why there is no significant change after removing those features. We used the best model to conduct the real-time knowledge gain prediction to find out how quickly we could predict the knowledge gain. According to the experiment results, we found the accuracy of prediction only had a very small decline (2.6%) from the first page to the last, and the biggest fluctuation was 3.3%. Given that the reading exercise lasted 30 min, the cause could be variations in the participant's reading pattern over time as a result of tiredness. In general, the prediction accuracy is fairly stable. We believe this result showed our model was able to make predictions on participants' knowledge gains from the very beginning.

Can we generalize our results to other scenarios?

In this study, we found that there were significant differences among the knowledge gain levels across scanning strategies and extensive and intense reading distributions, and that the robot's feedback had a significant impact on the entire reading process. Based on these findings, we built a model to predict the knowledge gain level. Our findings generalizability can be demonstrated by three factors, the first of which is the diversity of the gaze data. All participants were recruited at random from a campus population, and they all had various demographic profiles. Second, we compared and analyzed our results on the correlation between participant performance on the e-reading task and gaze features. Despite the impact of the different experimental settings, the majority of our results are in line with prior studies of a similar kind. Third, we established that our models should generalize to new users in similar situations by ensuring participant independence across training and testing sets.

Aspects of our results may also be generalizable to many settings for interaction, such as browsing websites or understanding program code. The 37 characteristics may all be assessed regardless of the content being examined, including fixation number and duration, saccade amplitude, and saccade angle. These features are very likely to generalize to a wide range of environments. Therefore, a far wider range of systems than those that show text, such as online lectures or information visualization systems, may be applicable to our findings. Of course, this is a hypothetical assertion that has to be supported by empirical evidence.

The generalizability of our findings was also constrained. First, methodological decisions limit the applicability of our study since the eye tracking devices, Pupil Core, utilized in our study restrict the use of eyeglasses and are prohibitively expensive (€3,440), making them unlikely to be widely used. The fact that the data were collected in a lab environment is a second factor that could limit how generally applicable our models are.

For future researchers with similar interests. Our findings show that "repair strategies" applied in the reading comprehension test benefit learner performance, and robot intervention that encourages rereading behavior may have a beneficial impact on knowledge development. As a result, in the future, researchers might experiment with alternative methods to successfully encourage learners to reread. It is also worth noting the significant correlation between knowledge level and the F-shaped scanning method. When researchers build the computerized reading interface, they might base the text arrangement on the F-shaped scanning habits of the learners.

8 Conclusion

In the present paper, we investigate how different reading strategies relate to knowledge gain. We discovered significant differences between different knowledge gain levels across scanning strategies and extensive and intensive reading distributions. The high knowledge gain group uses an F-shaped scanning method, focusing solely on the most relevant areas, which implies they read more extensively rather than intensively, which is indicated by fewer and shorter fixations, a larger saccade amplitude, and a higher horizontal saccade proportion. The low knowledge gain group, on the other hand, read more extensively across the text and scanned between lines and sections more frequently. Interestingly, there are no statistically significant differences between the two knowledge gain groups, as revealed by the reading pattern analysis on the AOIs. We discovered that the robot's feedback had a considerable impact on the entire reading process by comparing the participants' gaze features under different conditions. In the HUM condition, participants had wider gaze coverage, more and longer general fixes, and more and longer regressive fixes. We came to the conclusion that participants were more task-focused and that they reread the text when they felt their understanding was inadequate. We could automat-

ically distinguish two knowledge gain levels based on the feature analysis with an average accuracy of 74.2% for a soft voting classifier, which was an ensemble of NuSVC, linear discriminant analysis, k-nearest neighbors, and a Gaussian process classifier. This model also achieved stable knowledge gain level prediction accuracy in the real-time simulation setting, and the final accuracy was 71.5%. This work is a first step towards the design of an intelligent tutoring system that is able to provide adaptable and immediate feedback on participants' reading patterns and estimate their real-time knowledge gain during reading. Future research will include investigating the usability of additional modalities, such as body posture and facial expression, to predict knowledge gain.

Acknowledgements The data collection was supported by the Interactive Intelligence Group and the Web Information Systems Group at TU Delft. Additionally, the authors would like to thank Yoon Lee, Catharine Oertel, Marcus Specht, and Jennifer Olsen for their help with the work.

Author Contributions XL designed and performed the experiments, derived the models and analysed the data. JM, QW, and XL interpreted the data analysis results and wrote the manuscript together. All authors of this paper have read and approved the final version submitted.

Funding The first author is funded by the China Scholarship Council (CSC) (No. 202006120103) from the Ministry of Education of the P.R. China. This work was supported by the National Natural Science Foundation of China under Grant 61876054. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Availability of data and materials The eye gaze data that support the findings of this study are available from the Interactive Intelligence group of TU delft. The data may be available upon request but not for all due to relevant data protection laws.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix A: Questionnaire

See Tables 4, 5 and 6.

Table 4 Pre-test question design

Index	Questions	Choice		
1	In order to design a product in line with eco-design principles, designers should consider extended product lifetime, durability, recyclability, and degradability	True	False	Don't know
2	In order to design a product that can be easily disassembled, designers need to disassemble the product themselves and understand the processes of shredding	True	False	Don't know
3	The objective of design for remanufacturing is to make products more durable and easy to disassemble	True	False	Don't know
4	The value of reuse of complex second-hand products is about 10 to 50 times more than their scrap value	True	False	Don't know
5	Considering the whole production chain is important because modern products are complex, and now it is required by regulation	True	False	Don't know
6	The Ecodesign of energy-related products is not voluntary	True	False	Don't know
7	The substitution of materials mitigates the procurement and sustainability issues linked to the supply and reduces the need for Critical Raw Materials (CRMs)	True	False	Don't know
8	It is important to make the right decisions during the first steps of the design process because most of the costs of the products developed are actually paid during these phases	True	False	Don't know
9	Products designers should make products that are easy to landfill	True	False	Don't know
10	Many ideas and concepts may be used during product design in order to improve the remanufacturing of products	True	False	Don't know
11	The first option in reducing wasted critical raw materials at a production facility is diminishing defects in the production process	True	False	Don't know
12	Product chain integration mainly considers the viewpoints of internal stakeholders involved in the product design	True	False	Don't know
13	The Waste Electrical and Electronic Equipment (WEEE) Directive is not a mandated policy in the product design process	True	False	Don't know
14	Substitution of material could be a good strategy to improve product's sustainability because it makes cheaper but technologically enhanced products	True	False	Don't know

Table 5 Pop-up question design

Index	Questions	Choice		
1	In order to design a product in line with eco-design principles, designers should consider extended product lifetime, durability, recyclability, and degradability	True	False	Don't know
2	In order to design a product that can be easily disassembled, designers need to disassemble the product themselves and understand the processes of shredding	True	False	Don't know
3	The objective of design for remanufacturing is to make products more durable and easy to disassemble	True	False	Don't know
4	The value of reuse of complex second-hand products is about 10 to 50 times more than their scrap value	True	False	Don't know
5	Considering the whole production chain is important because modern products are complex, and now it is required by regulation	True	False	Don't know
6	The Ecodesign of energy-related products is not voluntary	True	False	Don't know
7	The substitution of materials mitigates the procurement and sustainability issues linked to the supply and reduces the need for Critical Raw Materials (CRMs)	True	False	Don't know

Table 6 Post-test question design

Multiple answer question: select ALL correct answers
It is important to make the right decisions during the first steps of the design proces because most of the costs of the products developed are actually paid during these phases
Products designers should make products that are easy to landfill
Many ideas and concepts may be used during product design in order to improve the remanufacturing of products
The first option in reducing wasted critical raw materials at a production facility is diminishing defects in the production process
Product chain integration mainly considers the viewpoints of internal stakeholders involved in the product design
The Waste Electrical and Electronic Equipment (WEEE) Directive is not a mandated policy in the product design process
Substitution of material could be a good strategy to improve product's sustainability because it makes cheaper but technologically enhanced products

References

1. Abidin AZ (2020) Students reading comprehension through scanning technique. J Asian Multicult Res Educ Study 1(1):28–35. https://doi.org/10.47616/jamres.v1i1.13
2. Almasri A, Ahmed A, Almasri N et al (2019) Intelligent tutoring systems survey for the period 2000–2018. Int J Acad Eng Res
3. Anderson JR, Boyle CF, Reiser BJ (1985) Intelligent tutoring systems. Science 228(4698):456–462. https://doi.org/10.1126/science.228.4698.456
4. Asmawati A (2015) The effectiveness of skimming–scanning strategy in improving students' reading comprehension at the second grade of SMK Darussalam Makassar. Engl Teach Learn Res J ETERNAL 1(1):69–83. https://doi.org/10.24252/Eternal.V11.2015.A9
5. Bainbridge WA, Hart JW, Kim ES et al (2011) The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot 3(1):41–52. https://doi.org/10.1007/s12369-010-0082-7
6. Bamkin M, Goulding A, Maynard S (2013) The children sat and listened: storytelling on children's mobile libraries. N Rev Child Lit Librariansh 19(1):47–78. https://doi.org/10.1080/13614541.2013.755023
7. Belpaeme T, Kennedy J, Ramachandran A et al (2018) Social robots for education: a review. Sci Robot 3(21):eaat5954. https://doi.org/10.1126/scirobotics.aat5954
8. Bickmore T, Kimani E, Shamekhi A et al (2021) Virtual agents as supporting media for scientific presentations. J Multimodal User Interfaces 15:131–146. https://doi.org/10.1007/s12193-020-00350-y
9. Biedert R, Hees J, Dengel A et al (2012) A robust realtime reading-skimming classifier. In: Proceedings of the symposium on eye tracking research and applications, pp 123–130. https://doi.org/10.1145/2168556.2168575
10. Biswas G, Leelawong K, Schwartz D et al (2005) Learning by teaching: a new agent paradigm for educational software. Appl Artif Intell 19(3–4):363–392. https://doi.org/10.1080/08839510590910200
11. Bixler R, D'Mello S (2016) Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Model User Adapt Int 26(1):33–68. https://doi.org/10.1007/s11257-015-9167-1
12. Block E (1986) The comprehension strategies of second language readers. TESOL Q 20(3):463–494. https://doi.org/10.2307/3586295

13. Blok H, Oostdam R, Otter ME et al (2002) Computer-assisted instruction in support of beginning reading instruction: a review. Rev Educ Res 72(1):101–130. https://doi.org/10.3102/003465430720011
14. Brysbaert M, Mitchell DC (1996) Modifier attachment in sentence parsing: evidence from Dutch. Q J Exp Psychol Sect A 49(3):664–695. https://doi.org/10.1080/713755636
15. Coppi AE, Oertel C, Cattaneo A (2021) Effects of experts' annotations on fashion designers apprentices' gaze patterns and verbalisations. Vocat Learn 14(3):511–531. https://doi.org/10.1007/s12186-021-09270-8
16. Crossley SA, Varner LK, Roscoe RD et al (2013) Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In: International conference on artificial intelligence in education. Springer, pp 269–278. https://doi.org/10.1007/978-3-642-39112-5_28
17. Davison DP, Wijnen FM, Charisi V et al (2021) Words of encouragement: how praise delivered by a social robot changes children's mindset for learning. J Multimodal User Interfaces 15:61–76. https://doi.org/10.1007/s12193-020-00353-9
18. Bixler ER, D'Mello KS (2021) Crossed eyes: domain adaptation for gaze-based mind wandering models. Association for Computing Machinery, New York, NY, USA, ETRA '21 full papers. https://doi.org/10.1145/3448017.3457386
19. Faber M, Bixler R, D'Mello SK (2018) An automated behavioral measure of mind wandering during computerized reading. Behav Res Methods 50(1):134–150. https://doi.org/10.3758/s13428-017-0857-y
20. Feng S, D'Mello S, Graesser AC (2013) Mind wandering while reading easy and difficult texts. Psychon Bull Rev 20(3):586–592. https://doi.org/10.3758/s13423-012-0367-y
21. Follmer DJ (2018) Executive function and reading comprehension: a meta-analytic review. Educ Psychol 53(1):42–60. https://doi.org/10.1080/00461520.2017.1309295
22. Garner R (1987) Metacognition and reading comprehension. Ablex Publishing, New York
23. Gernsbacher MA, McKinney VM (1999) Comprehension: a paradigm for cognition. Am Sci 87(6):568
24. Van der Gijp A, Ravesloot C, Jarodzka H et al (2017) How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ 22(3):765–787. https://doi.org/10.1007/s10459-016-9698-1
25. Golinkoff RM (1975) A comparison of reading comprehension processes in good and poor comprehenders. Read Res Q 11:623–659. https://doi.org/10.2307/747459
26. Gordon G, Breazeal C (2015) Bayesian active learning-based robot tutor for children's word-reading skills. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v29i1.9376
27. Guthrie JT, Wigfield A, Barbosa P et al (2004) Increasing reading comprehension and engagement through concept-oriented reading instruction. J Educ Psychol 96(3):403. https://doi.org/10.1037/0022-0663.96.3.403
28. Harmer J (2001) The practice of English language teaching. London/New York, pp 401–405
29. Hedge T (2001) Teaching and learning in the language classroom, vol 106. Oxford University Press, Oxford
30. Holmqvist K, Nyström M, Andersson R et al (2011) Eye tracking: a comprehensive guide to methods and measures. Oxford University Press, Oxford
31. Huang X, Craig SD, Xie J et al (2016) Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. Learn Ind Differ 47:258–265. https://doi.org/10.1016/j.lindif.2016.01.012
32. Hutt S, Krasich K, Mills C et al (2019) Automated gaze-based mind wandering detection during computerized learning in classrooms. User Model User Adapt Int 29(4):821–867. https://doi.org/10.1007/s11257-019-09228-5
33. Hyönä J, Lorch Jr RF, Rinck M (2003) Eye movement measures to study global text processing. In: The mind's eye. Elsevier, pp 313–334. https://doi.org/10.1016/B978-044451020-4/50018-9
34. Ince G, Yorganci R, Ozkul A et al (2021) An audiovisual interface-based drumming system for multimodal human-robot interaction. J Multimodal User Interfaces 15:413–428. https://doi.org/10.1007/s12193-020-00352-w
35. Kaakinen JK, Hyona J (2005) Perspective effects on expository text comprehension: evidence from think-aloud protocols, eyetracking, and recall. Discourse Process 40(3):239–257. https://doi.org/10.1207/s15326950dp4003_4
36. Kaakinen JK, Hyönä J (2007) Perspective effects in repeated reading: an eye movement study. Mem Cogn 35(6):1323–1336. https://doi.org/10.3758/BF03193604
37. Kaakinen JK, Hyönä J (2010) Task effects on eye movements during reading. J Exp Psychol Learn Mem Cogn 36(6):1561. https://doi.org/10.1037/a0020693
38. Kaakinen JK, Hyönä J, Keenan JM (2002) Perspective effects on online text processing. Discourse Process 33(2):159–173. https://doi.org/10.1207/S15326950DP3302_03
39. Kaakinen JK, Hyönä J, Keenan JM (2003) How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. J Exp Psychol Learn Mem Cogn 29(3):447–457. https://doi.org/10.1037/0278-7393.29.3.447
40. Kaakinen JK, Olkoniemi H, Kinnari T et al (2014) Processing of written irony: an eye movement study. Discourse Process 51(4):287–311. https://doi.org/10.1080/0163853X.2013.870024
41. Kassner M, Patera W, Bulling A (2014) Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In: Adjunct proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. ACM, New York, NY, USA, UbiComp '14 Adjunct, pp 1151–1160. https://doi.org/10.1145/2638728.2641695
42. Kennedy J, Baxter P, Senft E et al (2016) Social robot tutoring for child second language learning. In: 2016 11th ACM/IEEE international conference on human-robot interaction (HRI), pp 231–238. https://doi.org/10.1109/HRI.2016.7451757
43. Khachatryan GA, Romashov AV, Khachatryan AR et al (2014) Reasoning mind Genie 2: an intelligent tutoring system as a vehicle for international transfer of instructional methods in mathematics. Int J Artif Intell Educ 24(3):333–382. https://doi.org/10.1007/s40593-014-0019-7
44. Kidd CD, Breazeal C (2007) A robotic weight loss coach. In: Proceedings of the national conference on artificial intelligence. London, AAAI Press, MIT Press, Menlo Park, Cambridge, p 1985
45. Kontogiorgos D, Pereira A, Gustafson J (2021) Grounding behaviours with conversational interfaces: effects of embodiment and failures. J Multimodal User Interfaces 15:239–254. https://doi.org/10.1007/s12193-021-00366-y
46. Lee Y, Chen H, Zhao G et al (2022) Wedar: webcam-based attention analysis via attention regulator behavior recognition with a novel e-reading dataset. In: Proceedings of the 2022 international conference on multimodal interaction. Association for Computing Machinery, New York, NY, USA, ICMI '22, pp 319–328. https://doi.org/10.1145/3536221.3556619
47. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308. https://doi.org/10.1007/s12369-013-0178-y
48. Leite I, Pereira A, Mascarenhas S et al (2013) The influence of empathy in human-robot relations. Int J Hum Comput Stud 71(3):250–260. https://doi.org/10.1016/j.ijhcs.2012.09.005
49. Leyzberg D, Spaulding S, Toneva M et al (2012) The physical presence of a robot tutor increases cognitive learning gains. In:

Proceedings of the annual meeting of the cognitive science society. https://escholarship.org/uc/item/7ck0p200
50. Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum Comput Stud 77:23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001
51. Lin W, Yueh HP, Wu HY et al (2014) Developing a service robot for a children's library: a design-based research approach. J Am Soc Inf Sci 65(2):290–301. https://doi.org/10.1002/asi.22975
52. Michaelis JE, Mutlu B (2018) Reading socially: transforming the in-home reading experience with a learning-companion robot. Sci Robot 3(21):eaat5999. https://doi.org/10.1126/scirobotics.aat5999
53. Mousavinasab E, Zarifsanaiey N, NiakanKalhori RS et al (2018) Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interact Learn Environ 29(1):142–163. https://doi.org/10.1080/10494820.2018.1558257
54. Murray WS (2000) Commentary on section 4. Sentence processing: issues and measures. In: Kennedy A, Radach R, Heller D et al (eds) Reading as a perceptual process. North-Holland/Elsevier Science Publishers, pp 649–664. https://doi.org/10.1016/B978-008043642-5/50030-9
55. National Reading Panel (US) NIoCH, (US) HD (2000) Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. National Institute of Child Health and Human Development, National Institutes of Health
56. Oertel C, Coppi A, Olsen JK, et al (2019) On the use of gaze as a measure for performance in a visual exploration task. In: European conference on technology enhanced learning. Springer, pp 386–395. https://doi.org/10.1007/978-3-030-29736-7_29
57. Olson E (2011) AprilTag: a robust and flexible visual fiducial system. In: 2011 IEEE international conference on robotics and automation. IEEE, pp 3400–3407. https://doi.org/10.1109/ICRA.2011.5979561
58. Paris SG, Cross DR, Lipson MY (1984) Informed strategies for learning: a program to improve children's reading awareness and comprehension. J Educ Psychol 76(6):1239
59. Paris SG, Wasik B, Turner JC (1991) The development of strategic readers. In: Barr R, Kamil ML, Mosenthal PB et al (eds) Handbook of reading research, vol 2. Lawrence Erlbaum Associates Inc, Mahwah, pp 609–640
60. Pernice K, Whitenton K, Nielsen J et al (2014) How people read online: the eyetracking evidence. Nielsen Norman Group, Fremont
61. Pollard-Durodola SD, Gonzalez JE, Simmons DC et al (2011) The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. Except Child 77(2):161–183. https://doi.org/10.1177/001440291107700202
62. Pourhosein Gilakjani A, Sabouri NB (2016) How can students improve their reading comprehension skill. J Stud Educ 6(2):229–240. https://doi.org/10.5296/jse.v6i2.9201
63. Pressley M, El-Dinary PB, Brown R (1992) Skilled and not-so-skilled reading: Good information processing and not-so-good information processing. In: Pressley M, Harris KR, Guthrie JT (eds) Promoting academic competence and literacy in school. Academic Press, pp 91–127
64. Rajendran R, Kumar A, Carter KE et al (2018) Predicting learning by analyzing eye-gaze data of reading behavior. Int Educ Data Min Soc 16–20. https://api.semanticscholar.org/CorpusID:52173770
65. Raney GE, Campbell SJ, Bovee JC (2014) Using eye movements to evaluate the cognitive processes involved in text comprehension. J Vis Exp JoVE 83:e50,780. https://doi.org/10.3791/50780
66. Reichle ED, Pollatsek A, Fisher DL et al (1998) Toward a model of eye movement control in reading. Psychol Rev 105(1):125–157. https://doi.org/10.1037/0033-295X.105.1.125
67. Reynolds RE (2000) Attentional resource emancipation: toward understanding the interaction of word identification and comprehension processes in reading. Sci Stud Read 4(3):169–195. https://doi.org/10.1207/S1532799XSSR0403_1
68. Rumpf C, Boronczyk F, Breuer C (2020) Predicting consumer gaze hits: a simulation model of visual attention to dynamic marketing stimuli. J Bus Res 111:208–217. https://doi.org/10.1016/j.jbusres.2019.03.034
69. Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36(8):1627–1639
70. Sevcenko N, Appel T, Ninaus M et al (2023) Theory-based approach for assessing cognitive load during time-critical resource-managing human-computer interactions: an eye-tracking study. J Multimodal User Interfaces 17(1):1–19. https://doi.org/10.1007/s12193-022-00398-y
71. Shrestha S, Lenz K, Chaparro B et al (2007) "f" pattern scanning of text and images in web pages. In: Proceedings of the human factors and ergonomics society annual meeting. SAGE Publications, Los Angeles, pp 1200–1204. https://doi.org/10.1177/154193120705101831
72. Steenbergen-Hu S, Cooper H (2013) A meta-analysis of the effectiveness of intelligent tutoring systems on k-12 students' mathematical learning. J Educ Psychol 105(4):970–987. https://doi.org/10.1037/a0032447
73. Sugimoto M (2011) A mobile mixed-reality environment for children's storytelling using a handheld projector and a robot. IEEE Trans Learn Technol 4(3):249–260. https://doi.org/10.1109/TLT.2011.13
74. Van Dijk TA, Kintsch W et al (1983) Strategies of discourse comprehension. Psychology 6(6):12
75. Wang YH, Young SSC, Jang JSR (2013) Using tangible companions for enhancing learning English conversation. J Educ Technol Soc 16(2):296–309
76. Weinfurt KP (1995) Multivariate analysis of variance. In: Grimm LG, Yarnold PR (eds) Reading and understanding multivariate statistics. American Psychological Association, pp 245–276
77. Weisstein EW (2004) Bonferroni correction. https://mathworld.wolfram.com/BonferroniCorrection.html
78. Wijekumar K, Meyer BJ, Lei P et al (2020) Supplementing teacher knowledge using web-based intelligent tutoring system for the text structure strategy to improve content area reading comprehension with fourth-and fifth-grade struggling readers. Dyslexia 26(2):120–136. https://doi.org/10.1002/dys.1634

79. Yadollahi E, Johal W, Paiva A et al (2018) When deictic gestures in a robot can harm child-robot collaboration. In: Proceedings of the 17th ACM conference on interaction design and children, pp 195–206. https://doi.org/10.1145/3202185.3202743
80. Yang W, Dai W, Gao L (2012) Intensive reading and necessity to integrate learning strategies. Engl Lang Lit 2(1):55–63. https://doi.org/10.5539/ells.v2n1p112
81. Yueh HP, Lin W, Wang SC et al (2020) Reading with robot and human companions in library literacy activities: a comparison study. Br J Edu Technol 51(5):1884–1900. https://doi.org/10.1111/bjet.13016
82. Zhao Q, Yuan X, Tu D et al (2015) Eye moving behaviors identification for gaze tracking interaction. J Multimodal User Interfaces 9:89–104. https://doi.org/10.1007/s12193-014-0171-2
83. Zwaan RA, Singer M (2003) Text comprehension. In: Graesser AC, Gernsbacher MA, Goldman SR (eds) Handbook of discourse processes. Lawrence Erlbaum Associates Publishers, Mahwah, pp 83–121

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.