Affective and Pleasurable Design, Vol. 41, 2022, 224–231
https://doi.org/10.54941/ahfe1001801

Confidence Estimation in Multiple Choice Questions Using Eye Movements

Naoya Hamadate[1], Keiichi Watanuki[2], and Kazunori Kaede[2]

[1]Graduate School of Science and Engineering, Saitama University, 255 Shimo-okubo, Sakura-ku, Saitama-shi, Saitama 338-8570, Japan
[2]Advanced Institute of Innovative Technology, Saitama University, 255 Shimo-okubo, Sakura-ku, Saitama-shi, Saitama 338-8570, Japan

ABSTRACT

The purpose of this study is to quantitatively evaluate the relationship between eye movement data and the subjective evaluation of the confidence of a learner using multiple-choice questions that are often used in e-learning. It was assumed that their confidence could be estimated from their eye movements, such as fixations, saccades, and changes in their pupil diameters. The experiment was conducted using 12 subjects and a desktop eye movement measurement device. The results showed that the fixation duration of the subjects for the word questions, and the saccade gain and vertical saccades of them for the listening questions varied with their level of confidence in answering the questions, suggesting that the effective evaluation index changed depending on the type of question.

Keywords: Eye movement, Fixation, Saccade, Pupil diameter, Multiple-choice questions

INTRODUCTION

In recent years, the widespread use of electronic devices has promoted the adoption of e-learning as a common form of learning, particularly among learners and other young people. Although there are various types of e-learning questions such as essay, fill-in-the-blank, or multiple-choice, the grading of e-learning questions is almost always done automatically by a computer, limiting the use of written and fill-in-the-blank questions, which usually have subjectively correct answers. However, multiple-choice questions are often used because of their scoring accuracy and cognitive nature. However, using such questions is disadvantageous in that learners can answer correctly by guesswork or chance, without providing answers based on the content that should have been reviewed. Therefore, it is important to use data other than the amount of correct answers to estimate a learner's level of confidence in answering questions from their mental state and reflect it in their comprehension status.

Studies on using eye movements to understand learners' mental states have been conducted. A previous study (Kojima, 2014) reported that the more confident a learner was in answering a multiple-choice question, the greater their fixation was toward their perceived correct choice. Another typical eye movement is the saccade, which is a rapid eye movement that occurs when

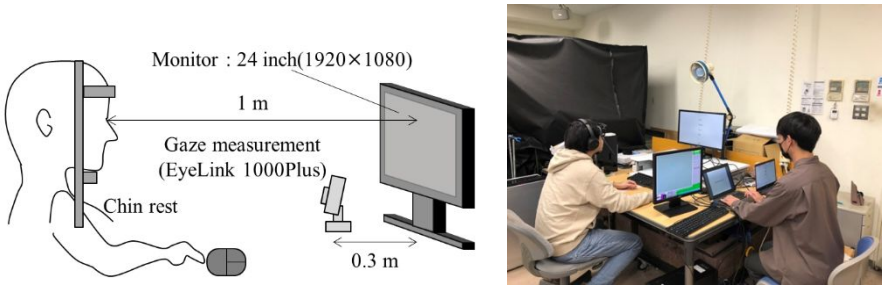

Figure 1: Experimental environment.

the target of someone's gaze is changed. It has been suggested that saccades are effective for assessing comprehension during text reading (Seo, 2002). In addition, some studies have focused on the pupil diameter, suggesting that it changes with variations in the concentration level of the learner during learning (Noma, 2011) and that this change is effective in evaluating the quality of their English word acquisition (Hirata, 2020). The present study aims to quantitatively evaluate the relationship between eye movement data and the subjective evaluation of a learner's confidence, assuming that confidence can be estimated from the fixations, saccades, and changes in the pupil diameter of the learner.

EXPERIMENT FOR COMPARING EYE MOVEMENTS

To evaluate the relationship between the confidence level and eye movements of a learner when answering multiple-choice questions, we prepared two types of four-answer choice questions and conducted a comparison experiment on eye movements using a subjective-confidence-evaluation questionnaire.

Experimental Method

The experimental setup is shown in Figure 1. The EyeLink 1000 PLUS, a desktop eye-movement measurement device, was used. The sampling frequency was set to 1000 Hz, and the device was placed 0.3 m away from the monitor. The experimental subjects were seated 1 m from the monitor, and their heads were fixed on a chin rest so that they directly faced the monitor.

The two types of four-answer choice questions were word questions (Japanese translations of English words) and English listening questions (Figure 2). For the listening questions, different words were explained by an English-speaking voice , and the subjects were asked to choose the appropriate response from the four displayed English words. For both types of questions, 30 were selected from among those prepared so that the level of difficulty would suitably vary.

In the experiment, 12 Japanese males (22.7±1.5 years old) were each given a word question and listening question. Then, 30 trials were conducted in random order. The flow of each trial is shown below (see Figure 3). First, the subjects were presented with the question. In the word question, an English

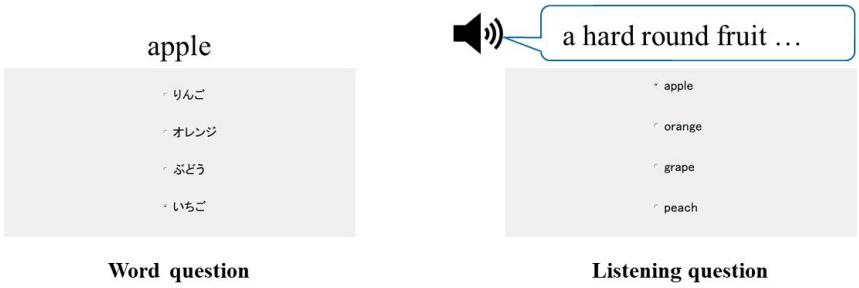

Figure 2: Details of the questions (word and listening questions).

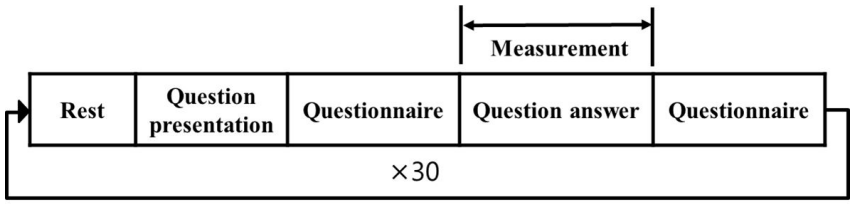

Figure 3: Flow of a single trial.

word was displayed on the monitor for 5 s, and in the listening question, English audio was inputted through a headset. A subjective evaluation questionnaire was administered to the subjects before and after they answered the questions. In the pre-questionnaire, they were asked to choose one of the following responses based only on their understanding of the question: "I know the answer," "I probably know the answer," "I probably don't know the answer," or "I don't know the answer at all." After they responded, the subjects were shown four choices and asked to answer the question immediately. Finally, in the post-questionnaire, they were asked to choose one of the following: "I could answer the question even without choices," "I knew the answer by looking at the choices," "I didn't know the answer, but I guessed from the choices," or "I didn't know the answer at all." After resting for 5 s, the next question would then be presented. The eye movements of the subjects were measured while they answered the questions.

Analysis Method

First, we explain how the fixation of each participant was evaluated. A region was set for each option based on the position of the screen while the participant answered, and the fixation time in the region was calculated for each of the four options (see Figure 4). The fixation time in the area of each of the four options was calculated, and an evaluation index for the fixation time ratio was obtained by dividing the fixation time in the area of the answered option by the total fixation time.

The evaluation indices for the saccades were saccade frequency, saccade gain, and the percentage of vertical saccades. The saccade frequency was calculated by dividing the number of saccades in the solution time by the solution time, and the saccade gain was calculated by dividing the average saccade gain by the solution time. According to a previous study (Shigeta,

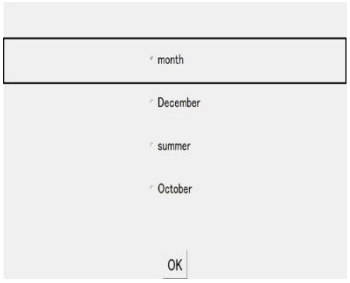

Figure 4: Definition range of the fixation point area.

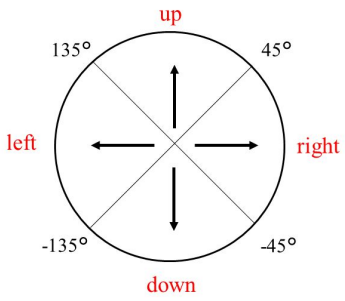

Figure 5: Definition of saccade direction.

2011), a person's gaze tended to move in different directions (horizontal and vertical) for multiple-choice questions of high difficulty. Therefore, we focused on the direction of these saccades. The horizontal right direction was set to 0° and the direction of the saccades from 45 to 135° was set in the upward direction. The saccades were then divided into four directions: left, down, left, and right in 90° counterclockwise units. In this experiment, all the options were aligned vertically and comprised only words, so the saccades generated when comparing the options were mostly vertical. Therefore, it could be inferred that the higher the confidence level, the higher was the percentage of vertical saccades that occurred. Accordingly, the percentage of vertical saccades was calculated by dividing the number of saccades in the upward and downward directions by the total number of saccades and it was used as an evaluation index.

Some data in the measured pupil diameter data were lost due to the blinking in the eyes of the subjects. Thus, the pupil diameter data needed to be corrected. It has been suggested that the range of the blinking can be determined from the corresponding fluctuations in the pupil diameter data and the differences in the eyelid movements. Therefore, we defined the blinking rate using the blink detection method described in a previous study (Hershman, 2018). Pupil diameter data in the blink range were removed and linearly interpolated. Since the pupil diameter underwent random temporal variations (Mathôt, 2018), the average pupil diameter was evaluated by subtracting the average pupil diameter during the first 10 ms after the start of the solution time.

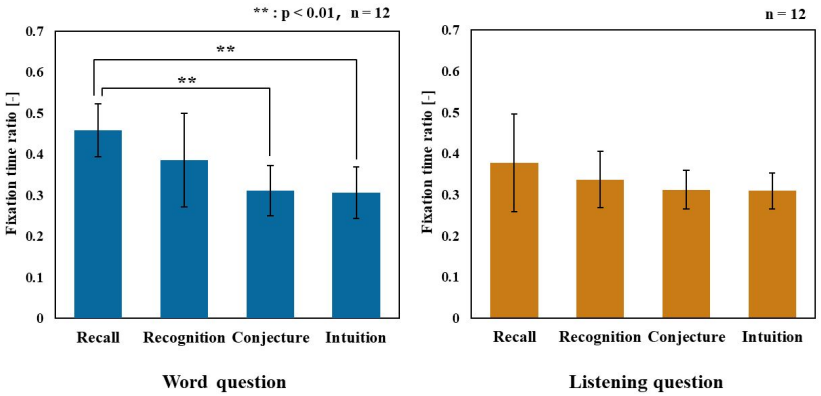

Figure 6: Results for fixation time ratios.

The confidence level of the participants was classified into four groups based on their responses to the post-intervention questionnaire in order of increasing confidence: "Recall," "Recognition," "Conjecture," and "Intuition." The mean values of the evaluation indices of each group were calculated for each subject and they were compared among the four groups.

RESULTS

The differences in the means of the evaluation indices were compared using the Steel-Dwass test. The results for the fixation time ratios are shown in Figure 6. For the word questions, the "Recall" group was significantly larger than the "Conjecture" group, and the "Recall" group was significantly larger than the "Intuition" group at the 1% significance level, but there was no significant difference for the listening question.

The results for the saccades were as follows. First, there was no significant difference in saccade frequency for either type of question (see Figure 7). For the word questions, there was no significant difference in the saccade gain or vertical percentage, whereas such a difference was observed in for the listening questions, with "Recall" being significantly larger than "Conjecture" and "Intuition" at the 5% and 1% significance levels, respectively. In addition, "Recognition" was significantly larger than "Intuition" in the saccade gain and vertical saccade ratio at the 1% and 5% significance levels, respectively. (see Figures 8 and 9).

The results of the average pupil diameter were not significantly different for either the word or listening questions (see Figure 10).

DISCUSSION

The results showed that there was a significant difference between the two types of questions, which was due to the difference in the fixation time for the word questions and in the saccade for the listening questions according to the confidence level of the subject. This difference may be due to the variations in the cognitive load based on the differences in the type of questions that

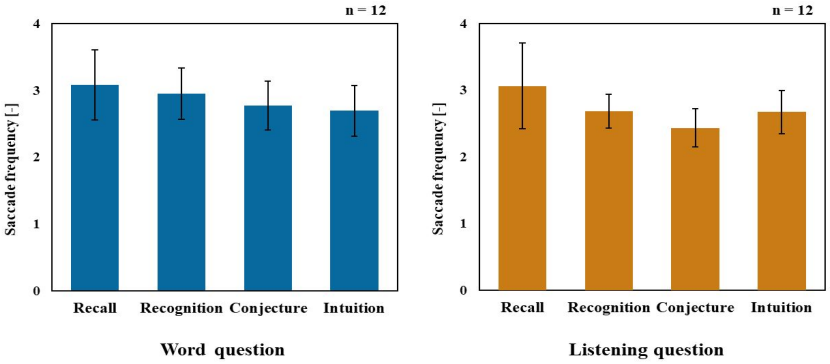

Figure 7: Results of saccade frequency.

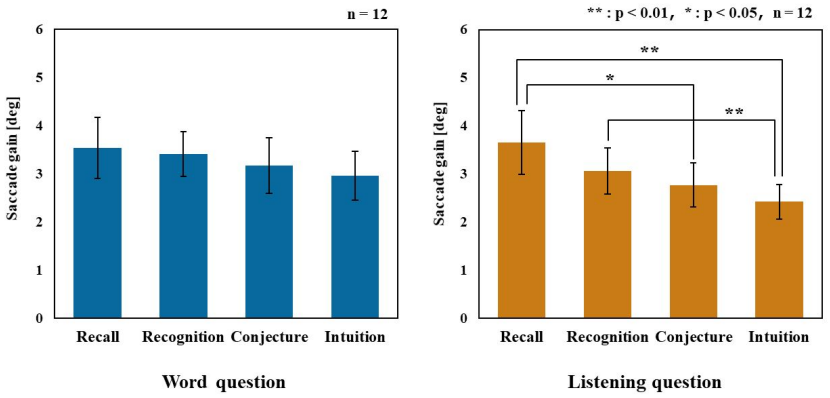

Figure 8: Results of average saccade gain.

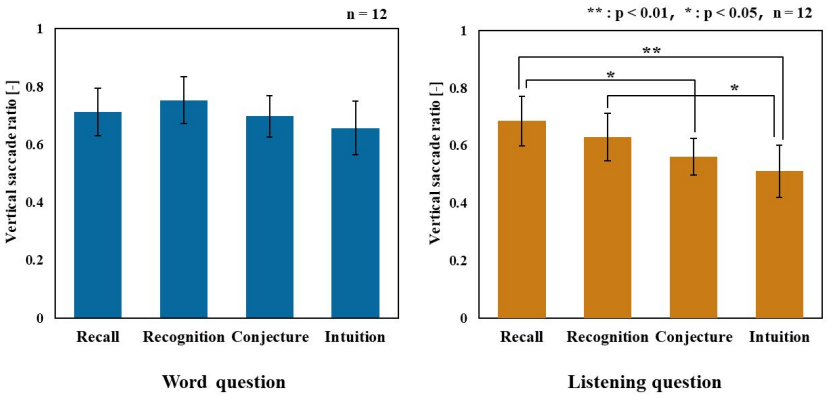

Figure 9: Results of vertical saccade ratios.

were answered. The word questions were simpler and required less cognitive load because they could be solved using only knowledge of the word. On the other hand, the listening questions were presented as sentences and required memory retention till the question had to be answered; therefore, the cognitive load was high.

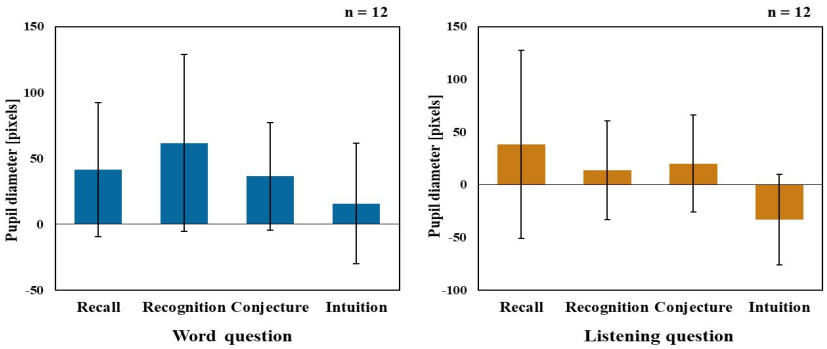

Figure 10: Results of average pupil diameters.

First, we considered the fixation time. In the word questions, it was assumed that in confident questions, the answer was given by fixation in almost the entire area after selecting an option, whereas in unconfident questions, the answer was given by comparing the options and selecting the one that seemed most appropriate. Therefore, the lower the level of confidence, the poorer and longer the fixation. However, in the listening task, there were eye movements, comparing the alternative options, even for questions for which the learner was certain of the answer; this was not seen as a change in fixation time.

As for the saccades, it was assumed that eye movements during "swimmer's eye," caused by hesitation and agitation when the confidence level of the subject was low, were reflected in the saccade gain and the ratio of vertical saccades in this experiment. For listening questions, because of the large amount of thinking required before answering the questions, hesitation and frustration appeared as significant changes, and differences in the saccades were observed. On the other hand, for word questions, the learners answered only with their knowledge and did not think much, so the phenomenon of "swimmer's eye" did not occur and no differences in saccades were observed.

CONCLUSION

The purpose of this study is to quantitatively evaluate the relationship between data on eye movements and the subjective evaluation of confidence in individuals while answering multiple-choice questions. We conducted experiments to observe the eye movements of subjects while they answered two types of multiple-choice questions. We found that fixation time, saccade gain, and saccade direction were effective for quantitative evaluation. It was also suggested that the characteristics of the eye movements differed depending on the amount of cognitive load and that the effect of each evaluation index for each type of question on estimating the confidence of individuals should be considered. In the future, we consider creating a model that allows practical estimation of the confidence level of individuals using the characteristics of their eye movements and conducting a regression analysis for it.

REFERENCES

Hershman, R., Henik, A., Cohen, N. (2018). A novel blink detection method based on pupillometry noise", Behavior research methods, Volume 50 No. 1. pp. 107–114.

Hirata, T., Mishima, K., Fukaya, K., Hirata, Y. (2020). Pupillary responses as an objective measure of 2nd language words' comprehension for Japanese English learners, IEICE technical report, Volume 120 No. 185. pp. 87–92.

Kojima, K., Muramatsu, K., Matsui, T. (2014). Experimental Study on Description of Eye-Movements among Choices in Answering to Multiple-Choice Problems, Journal of Information and Systems in Education Volume 31 No. 2. pp. 197–202.

Mathôt, S., Fabius, J., Heusden, E.V., Stigchel, S.V. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data, Behavior research methods Volume 50 No. 1. pp. 94–106.

Noma, K., Ogawa, K. (2011). Patterning of the pupil response in learning support system on the calculation, IPSJ SIG Technical Report, Volume 2011-11-24 No. 1. pp. 1–6.

Seo, H., Lee, C. (2002). Head-free reading of horizontally and vertically arranged texts, Vision Research Volume 42 No. 10. pp. 1325–1337.

Shigeta, A., Koike, T., Hamamoto, K., Nosu, K. (2011). Estimation of Users' Subjective Difficulty of WEB Based English Listening Tests, IEEEJ Transactions on Electronics, Information, and Systems, Volume 131 No. 4. pp. 800–807.