Open Access





Juan Zhou^{1*}, Siqi Wang², Ling Xu³ and Chengjiu Yin⁴

*Correspondence: juan.z.kt@gmail.com

 ¹ School of Environment and Society, Tokyo Institute of Technology, CIC-809, 3-3-6 Shibaura, Minato-ku, Tokyo, Japan
 ² Graduate School of System Informatics, Kobe University, Kobe, Japan
 ³ Applied Information Technology, The Kyoto College of Graduate Studies for Informatics, Kyoto, Japan
 ⁴ Research Institute for Information Technology, Kyushu University, Fukuoka,

Japan

Abstract

Information avoidance has been studied in medicine, economics, and psychology, and has recently been discussed in educational technology. In this study, the authors developed a grouping method to reduce students' information avoidance in reading through group work. This two-step group method includes the k-means and genetic algorithm to explore the grouping method based on students' marking tendencies. To examine the effect of this method, an experiment was conducted in a web-system development course with 33 graduate students. The results showed that information avoidance occurred less in the experimental group than in the control group. The students of the two-step grouping method evaluated group work as more helpful for their study than the students who attended the usual group work.

Keywords: Information avoidance, Reading activity, Genetic algorithm, Graduate students

Mathematics Subject Classification: 0000, 1111

Introduction

Reading online has become an important learning method for college students. Students read academic literature, textbooks, and material from teachers to immerse in the discipline and gain knowledge (Hermida et al., 2009). More universities are increasingly experimenting with online study and e-books for instruction.

However, information avoidance (IA) occurs in reading behavior. IA is any behavior that prevents or delays information acquisition (Sweeny et al., 2010). It has different research directions in different contexts. Most people avoid information because they have negative emotions toward it; for example, they have different ideas about the views expressed in the information, or they are afraid of the implications of the information. Though IA is widely discussed in other fields, research on IA in the educational technology field, and especially its impact on reading behavior, is limited. Similar to other fields, students will subconsciously interpret information according to their cognition when



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/.

they hold opposing views on information, resulting in information loss. Students will skip selected parts of an article because of their resistance to the content of those parts. Furthermore, they may lose or misinterpret the information because they ignore certain parts of an article. This behavior significantly reduces the effectiveness of students' academic reading.

Fuertes (2020) suggested that IA had a positive correlation with reading strategies and attitude. Moreover, Hermida stated that college students must attain a certain reading proficiency before admission to support them in understanding the learning content (Hermida et al., 2009). However, students may lack sufficient ability to read literature; hence, they may lack a positive attitude toward reading. Therefore, IA will occur when they cannot reach the desired reading ability.

Hence, this study maintains that remedial methods are needed to help students alleviate consequences when encountering IA. This study proposes a method to support students in regaining the information lost in the reading process through group work based on their post-reading marking habits.

Literature review

Information avoidance

IA is discussed in the fields of psychology, economics, and physical health. Psychology research has shown that people avoid receiving information that conflicts with their worldview, called selective exposure (Covington & Mueller, 2001). In economics, people avoid information that makes them mentally uncomfortable or increases cognitive dissonance and uncertainty (Golman et al., 2017). People who refuse to accept physical health information become anxious, while people who actively receive health information improve their wellbeing (Ek & Heinström, 2011). Additionally, recent studies have discussed IA in order to understand health information behavior during a global health crisis (COVID-19) (Soroya et al., 2021).

IA is a phenomenon in which people cannot obtain information they deem unwanted (Sweeny et al., 2010). This includes information that they subjectively resist and cannot objectively accept. In reading, the most notable effect is that students skip content when reading the literature (Fuertes et al., 2020). When students have a positive attitude toward reading, they are more likely to employ better reading strategies and less likely to exhibit IA. This study clarifies the conditions under which students lose information after reading literature, based on their attitudes.

Regarding the causes of IA, there have been several summaries from different perspectives. Five reasons for IA, summarized by Golman et al. (2017), in line with reading behavior are physical avoidance, inattention, biased interpretation of information, forgetting, and self-handicapping. Physical avoidance occurs when students are reluctant to read articles, inattention and forgetting lead students to miss information, and biased interpretation of information and self-handicapping lead to misunderstanding of literature. Refusing to read literature is a problem of students' psychological state. This study explored a method to support students in obtaining information that they ignore while reading. Therefore, IA, in the scope of this study, occurs due to a combination of the above reasons.

Information avoidance and reading behavior

Some researchers have conducted experiments (Fuertes et al., 2020) on IA in academic reading. The experiments explored the influence of attitudes and reading strategies on IA. They concluded that students' reading attitudes and strategies positively impact IA. The more reading strategies are used, the lower the IA. Group study can effectively improve students' motivation (Maqtary et al., 2019) and provide a community environment for students to exchange information acquired from the literature.

Group work

In university education, group work is a common educational method, which aims to improve in-depth learning capabilities and cultivate teamwork skills. This study uses the method of group work to reduce IA. Discussions can allow students to exchange information that they consider important. Furthermore, it allows students to regain lost information due to IA.

Learning analytics (LA) refers to data analysis and interpretation related to learners' behaviors and interactions during the learning processes and their profiles and learning contexts (Gwo-Jen et al., 2017). Several researchers have reported that LA can be beneficial for different roles. Ren et al. (2017) suggested that research on reading logs could effectively promote students' reading outcomes (Ren et al. 2017). Therefore, this research focuses on word markings that could better reflect students' understanding of the literature. In group work, the group members should play different roles according to the group's mission and members' behaviors. Roles define how a person is expected to behave, contribute, and relate to others in collaborative work (Maqtary et al., 2019). In Chen et al. (2019) experiment, they positioned students' roles according to their communication tendencies.

Marking is a behavior that connects information and thinking in reading activities (Schilit et al., 1998). Some articles that are not marked may not necessarily indicate information evasion. Moreover, the marked sections indicate in-depth attention. This rationale underpins our decision to focus on marking behavior in our study. Hence, the work should be grouped to consider students' reading tendencies, which can be analyzed from the reading log data.

Method

Research purpose

This study aimed to develop a grouping method that considers students' reading tendencies to reduce their IA. By grouping students, the authors speculate that groups of students who avoid different parts of an article will exhibit significant knowledge differences. The more times students are exposed to content, the more likely they are to encounter information previously avoided. The data on students' marking habits can intuitively show their reading process. Therefore, this study examines whether grouping students according to their marking habits can effectively alleviate IA.

Information avoidance in reading behavior

Zhou and Yin (2023) defined three kinds of reading behavior states related to IA—excellent reading, skipped reading and missed reading (Fig. 1). This research focuses on missed reading and aspects of skipped reading.



Fig. 1 IA in reading behavior

Marking behavior

The markings that students make during the reading process can intuitively reflect their IA. There is a high probability that the content marked by the students has been seen and not been ignored. In addition, students' markings can reflect their reading emphasis. If a part of the article is heavily marked, the students likely paid more attention to its content. However, if a part is not marked, the student likely overlooked it. Therefore, students' marking habits reflect their IA. According to previous reading logs and observations, students' marking habits can be divided into four categories: high-frequency words, high-frequency sentences, low-frequency words, and low-frequency sentences. Furthermore, the marking categories can be bifurcated into two reading characters—the length of the markings and the time they were made (Fig. 2).

Two-step grouping method

The authors classified the different types of marking through the K-means algorithm. Subsequently, they selected students from each type through genetic algorithm. The classification processes were implemented in Python. As shown in Fig. 2, students were sorted and grouped in two steps. In the first step, students who marked similar words' lengths and similar marking frequencies were selected into the same groups. We used the K-means algorithm (Lloyd 1982) and set the group count to four. Variables for k-means are the length of words and times of the students marking them. After that, in the second step, students with similar reading rhythms were grouped. Page forward frequency and reading time were used for the genetic algorithm for the second step. The first classification homogeneously divided students into four marking types, and the second placed students with different marking types in the same group to assess the communication effect between groups.



Fig. 2 The illustration of two-step grouping method

Grouping by the k-means algorithm

K-means was used for clustering. Clustering centers on k points in space, and the objects closest to them are classified (MacQueen et al., 1967). Through an iterative method, the value of each cluster center is updated successively until the best clustering result is obtained. Applying k-means to this research, the parameters collected for student classification were the times of marking and the length of the words marked. In the coordinate system, with this parameter as the coordinate axis, students closest to each other are divided into the same class. In the calculation process, the formula for the distance between two points is as below.

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^{m} (X_{it} - C_{jt})^2}$$
(1)

In the formula, X is n different object points, that is, the marking parameters of students, and C is each cluster center obtained through each cycle. Computation ceases when the classification result no longer changes. The algorithm follows the following four steps.

- 1. Take K objects as the initial cluster centers.
- 2. Calculate the distance between each object and cluster center.
- 3. Assign each object to its nearest cluster center.
- 4. Recalculate cluster centers based on the existing objects in the cluster.

If the data of the total marking times and the total number of words marked by the students is used, an average number of marks per time would be obtained. However, this processing method would overlook substantial information. For example, if a student is accustomed to marking keywords and marks a long sentence at the end, the data of the word count of this long sentence will pollute the classification result of this student. To reduce the impact of this extreme marking phenomenon and classify students more accurately, the marking situation of each page was collected.

Grouping by genetic algorithm

Genetic algorithm is a computational model designed and proposed according to the evolutionary law of survival of the fittest in Darwin's theory of evolution (Mirjalili & Mirjalili 2019; Katoch et al., 2021). The process of solving the problem is converted into the process of crossover and mutation of chromosome genes in biological evolution.

After k-means classification, students were divided into four types, and each type was placed into different groups. The students with the same marking type shared the same reading personality. The reading personality identified that students shared the same tendency to avoid information during reading. For example, high-frequency sentence students marked more information than high-frequency word students. High-frequency word students possibly pay more attention to the keywords than the sentences. Highfrequency sentence students focus on the sentences rather than the words, meaning they likely notice more information but may miss important words. Hence, the second step of this method was designed to ensure that every group included different marker types. Different students in the same group communicate about their reading priorities and complement each other.

Considering the ease of group communication, students' reading time and page-turning frequency were variables. Both variables can reflect students' reading rhythm to a certain extent. In this manner, students have the same information exposure time, and the communication between them can be guaranteed to be fair. The authors avoided situations in which it was difficult to obtain valid information from students who had less time with the information. This type of consistent reading rhythm is called rhythm adaptation. The students were divided into different groups based on similar reading rhythms.

The genetic algorithm is a cycle algorithm, and the grouping results were generated after a set cycle. In the genetic algorithm, the judgment method is the most important aspect, called fitness value. To decrease the difference in reading parameters among the members of each group, the sum of the variances of each group is set as the fitness value. The formula for calculating the variance of group S1 is as follows.

$$S_1 = \frac{\sum_{i_1}^n (a_i - a)^2}{n} + \frac{\sum_{i_1}^n (b_i b)^2}{n} i_1 = 1, 2, 3, 4$$
(2)

where a is the reading time and b is the page-turning frequency. Adding the variance results for groups 1, 2, and 3 provides the total variance sum S. The lower the value of S, the better the result. Before evaluating the results, the number of marking habits in the group is determined. If more than half of the groups do not have four different types of marking habits, then a relatively large value will be added to the result to eliminate it.

Experiment

The experiment was carried out as a part of the class, in which 43 college students from two classes of the web-system development course joined the study. The names of the two classes were "System Design" and "Mobile Application Development". In the lectures for both of these classes, reading the Python textbook was included as part of the content. This experiment utilized class time for reading the Python textbook. We collocated



Fig. 3 E-book system interface

the valid data from a total of 33 college students, with 14 in the experimental group and 19 in the control group.¹ They were all master's students in the same major at a university, and they have some basic knowledge of computer science. The experiment was conducted during online classes using the Zoom platform. The content of the materials was extracted from a Python textbook designed for the class.

E-book system

The e-book platform used in this experiment was developed on the DITel platform. The DITel was designed by Yin et al. (2017), with pageturning, marking (highlight and underline) and note-taking functions. The DITel interface is presented in Fig. 3.

The logs collected and recorded include page-turning, dwell time, note contents, and other reading logs. Table 1 presents the example of a log record.

Experiment design

Student participants from two classes were assigned to the experiment or control groups. Each group was subjected to a preliminary and main experiment. The basic experiment information is presented in Table 2.

The preliminary experiment (Table 3) collected the marking log data of the students in the experimental group for grouping and helping both groups become familiar with

¹ The reasons for invalid data are repeated or unsubmitted tests or questionnaires.

Table 1	Example	e of marking	information	recorded by	y the e-boo	зk

User id	Operation name
python0	ADD HL id:100 page:0 bookpage:0 start: end: color: markertext: す. 織
python0	DELETE HL id:100 page:0 start: end: color: markertext: す. 織
python0	ADD HL id:100 page:0 bookpage:0 start: end: color: markertext:織物は
python0	DELETE HL id:100 page:0 start: end: color: markertext:織物は
python0	DELETE HL id:100 page:0 start: end: color: markertext:あると思
python0	ADD HL id:100 page:0 bookpage:0 start: end: color: markertext: が可能で す.
python0	DELETE HL id:100 page:0 start: end: color: markertext:が可能です
python0	ADD HL id:100 page:0 bookpage:0 start: end: color: markertext:ながら, 完成

Table 2 Basic information of the preliminary and the main experiment

Experiment name	Participating group	Date	The number of students
Preliminary experiment	Experimental group	2022.11.17	14
	Control group	2022.12.21	19
Main experiment	Experimental group	2022.12.22	14
	Control group	2022.12.21	19

Table 3 Experiment design

	Minutes	Graduate students of a university in Japan (N $=$ 33)				
		Experimental group (N $=$ 14)	Control group (N = 19)			
Preliminary experiment	10	Introduction to reading task, pre-test				
	15-20	Students read the materials using e-b	ook			
	10	Post-test and questionnaire				
Main experiment	10	Introduction to reading task, pre-test	and pre-questionnaire			
	15-20	Students read the materials using the	e-book			
	15-20	Discuss in a group which is grouping by the two-step grouping method	Discuss in a group which is grouping by random			
	10	Post-test and questionnaire				

the e-book system. The code for grouping and categorizing students was developed in Python. During the experiment, the reading logs of students' reading time, page-turning frequency, and marking content were used.

In the main experiment, students were divided into several smaller groups within both groups. The experimental group was grouped by the two-step grouping method, while the control group was grouped randomly. All the students discussed their readings within their groups. The main experiment explored students' IA based on the grouping method.

Evaluate information avoidance

In previous studies, IA was primarily evaluated through self-reporting (Fuertes 2020). In this study, the main purpose is to find solutions to reduce IA in reading and take a more

comprehensive approach by integrating self-reported data from the students, log data from the e-book, and questionnaire responses to discuss IA from various perspectives.

In this study, the evaluation standard of IA occurrence was designed according to the student's self-assessment in the post-test. After each question, students were asked about their answers. If they answered incorrectly, they were asked for the reason. If the student reported that they did not see the relevant content in the article or made a missed judgment, it was determined that the student had IA.

Test and questionnaire

During the experiment, the students answered two tests and two questionnaires before and after the experiment. The pre-test was used to assess students' level and evaluation of their reading situation, and the post-test was used to test students' learning achievement and evaluation of IA by themselves after their reading and group work. The prequestionnaire and post-questionnaires were used to assess the student's attitudes toward reading and group work. The tests and questionnaires were filled out by the students via Google Forms. The tests contained mostly multiple-choice questions, while the questionnaire contained mostly multiple-choice questions with Likert-scale responses. Examples of tests and questionnaires are as follows.

- Example of pre-test What is the computational setup method of the early computer (ENIAC)?
 - A. By changing the electronic component
 - B. By changing the hard disk
 - C. By changing the cable
 - D. I don't know
- Example of post-test Where can the result of the calculation be stored? A. Memory B. CPU C. Hard disk If your answer to this question is wrong, please explain why.
 - A. I don't think I did anything wrong
 - B. I didn't see it (the part related to the question)
 - C. Missed judgment
 - D. I forgot it
 - E. Other
- Example of Likert-scale question in the questionnaire. Are you good at reading? Please answer on a scale from 1 to 5 Yes 1---2--3--4--5 No

Data collection and analysis

The log data were collected through the e-book platform, including logid, courseno, coursecode, userno, userid, processcode, operationname, operationdate, ebookno, ebookid, ebookname, devicecode, deviceid, memo_text, page_no, scale, start_line, end_ line, pages, description, color, markertext, and type. All the test calculations satisfied



Fig. 4 The result of the clustering into four groups using k-means

the prerequisites, and data analysis was performed using the R.² We also used a missing value processing method to substitute some missing values in the questionnaire results with the average value.

Results

Preliminary experiment

As shown in Table 4, in the preliminary experiment, 437 codes of reading log data of the experimental group were successfully collected. Based on the data, the experimental group was divided into four groups according to the two-step grouping method, while the control group, which had more students, was divided into five groups randomly.

As Fig 4 showed, the spots with different colors represent the different types of students. In Fig. 5, the final result obtained from the system is shown. As described in the chapter 3.4.2, the lower the fitness value, the better the result. The wave shows the result of 500 iterations, and it can be seen that the lower fitness value shown around almost 100 times will be the best fitness value (- 92.05). The array of the best fitness value is shown below the picture.

Analysis of students' reading experience

Before the experiment, the results of students' reading experience showed no significant difference between the two groups (Table 5).

Analysis of information avoidance

In this study, two IA dimensions—skipped reading and missed reading—were measured. Table 6 illustrates the results of both dimensions. The results for both dimensions were higher for the control group than for the experimental group. Especially in the skipped reading dimension, there was a significant difference between the two groups (t = -2.24, p < 0.05).

² R Foundation. https://www.rproject.org.



Fig. 5 The result of fitness value obtained from genetic algorithm

Table 4 Grouping result

Log data	Students	Number of groups	Students in a group
437	14	24	3 or 4

 Table 5
 Descriptive data and t-test of students' experience and confidence results

Variable	N	Mean	SD	t value	df
Read articles					
Experimental group	14	2.71	0.91	0.09	30.12
Control group	19	2.68	1.06		
Proficient in reading articles					
Experimental group	14	2.36	1.15	- 0.61	22.64
Control group	19	2.58	0.84		

 Table 6
 The t-test results of skipped reading and missed reading for the two groups

Variable	N	Mean	SD	t value	df
Skipped					
Experimental group	14	0.57	0.76	- 2.24*	22.92
Control group	19	1.84	2.32		
Missed judgment					
Experimental group	14	0.21	0.43	- 2.06	25.9
Control group	19	0.74	0.99		

* *p* < 0.05

Variable	N	Mean	SD	t value	df
Pre-test					
Experimental group	14	4.78	1.31	- 3.23*	29.82
Control group	19	3.21	1.47		
Post-test					
Experimental group	14	6.64	1.28	- 3.54*	30.26
Control group	19	4.58	2.06		
Growth					
Experimental group	14	1.86	1.56	- 0.82	30.40
Control group	19	1.37	1.86		

Table 7	Descriptive	data and	t-test of the	pre-test/	post-test results
---------	-------------	----------	---------------	-----------	-------------------

* *p* < 0.05

Table o Palled t-lest of the pre-test/post-test results of the two group	lps
--	-----

Variable	N	Mean	SD	t value	df
Experimental-group					
Pre-test	14	4.78	1.31	- 4.45*	13
Post-test	14	6.64	1.28		
Control group					
Pre-test	19	3.21	1.47	- 3.20*	18
Post-test	19	4.58	2.06		

* *p* < 0.05

Fable 9 The t-test r	esults of	value of	group	discussion
----------------------	-----------	----------	-------	------------

Variable	N	Mean	SD	t value	df
Pre-questionnaire					
Experimental group	14	3.57	0.85	- 0.45	30.89
Control group	19	3.74	1.24		
Post-questionnaire					
Experimental group	14	3.71	0.83	- 2.61*	28.6
Control group	19	2.95	0.85		

* *p* < 0.05

Analysis of learning achievement

As presented in Table 7, there was a significant difference between the two groups before and after the experiment (t = -3.23, t = -3.54, p < 0.05). However, the experimental group scored higher than the control group in both scenarios. Moreover, the experimental group showed higher growth (1.86) than the control group (1.37). However, as shown in Table 8, there were significant differences between the pre-test and post-test in both groups (t = -3.20, t = -4.45, p < 0.05).

Analysis of group discussion satisfaction

Table 9 presents the t test results of the group discussion satisfaction of the two groups. There was a significant difference between the groups in the post-test

Variable	N	Mean	SD	t value	df
Pre-questionnaire					
Experimental group	14	3.21	0.97	0.14	30.89
Control group	19	3.16	1.26		
Post-questionnaire					
Experimental group	14	3.43	1.02	- 1.32	21.5
Control group	19	3.84	0.69		

Table 10 The t-test results of the value of reading attitude

Table 11 The t-test results of the value of marker and correct answer rate

Variable	N	Mean	SD	t value	df
Total of marked related test					
Experimental group	14	5.36	2.13	0.77	29.70
Control group	19	4.58	3.66		
Marker and answer correct					
Experimental group	14	3.07	2.37	- 3.08**	22.28
Control group	19	0.79	1.68		
No marker & answer correct					
Experimental group	14	2.43	1.87	- 1.04	28.51
Control group	19	1.74	1.91		

* *p* < 0.05, ** p < 0.01

questionnaire (t = -2.61, p < 0.05), while there was no significant difference between them in the pre-test questionnaire (t = -0.45, p > 0.05).

Analysis of reading attitude

Self-evaluation of the reading attitude was investigated through the questionnaire, and the results are presented in Table10. There was no significant difference between the results before (t = 0.14, p > 0.05) and after (t = -1.32, p > 0.05) the experiment.

Analysis of marker and correct answer rate

The content marked by each student was collected, and the marking content related to the question was extracted to judge whether each question was marked. Table 11 presents the number of marked questions and the number of marked questions answered correctly for both groups and the t-test results. There is no significant difference in the total number of questions marked between the two groups. More questions were marked and answered correctly in the experimental group than in the control group, with a significant difference (t = -3.08, p < 0.01).

Discussion and conclusions

In this study, a grouping system was designed to reduce students' IA through group discussions. This two-step (k-means and genetic algorithm) group method explored student groupings based on their marking habits. K-means divided students with the same marking habits, and the genetic algorithm divided students with different marking habits into the same group. Two experiments, including a preliminary and a main experiment, were conducted in a web-system development course. The results showed that compared to traditional grouping, the two-step grouping method significantly reduced students' IA occurrences. Compared with the control group, the number of times IA occurred in the experimental group decreased significantly. The students who went through the two-step grouping method evaluated the group work as more helpful for their study than the students who were randomly grouped.

A significant difference in learning performance was observed between the two groups before and after learning. Both groups received higher scores after learning; hence, the students who studied in the two-step group method were at par with the usual group. Moreover, there was no difference in students' study attitudes, reading experience, and confidence.

The experiment results confirmed that grouping students according to marking habits reduced IA and improved academic reading. It reduced the frequency of IA occurrence. Moreover, the experimental group evaluated group discussion effects more positively than the control group, although there was little difference between the two groups' learning performance and knowledge.

There was no difference in the number of notes students took between the experimental and control groups when reading books (related to test questions). However, many students answered correctly in the post-test after group discussion. The students found group discussions to be helpful. The students who participated in the two-step group work reinforced what they learned through group discussion. Even though there was no significant difference, the students in the two-step group answered more unlabeled questions correctly than the control group. Therefore, the study concludes that group learning through two-step grouping benefits students and reduces IA.

Comparison with previous studies

Compared with previous research on IA, this study starts from the data and studies IA according to the data characteristics. Previous research on IA has focused on psychological aspects. Research on academic reading, such as Fuertes' (2020) research, has been conducted using questionnaires and psychological research. Students have psychological IA during reading. However, the psychological factors of students' IA are complex, and different articles may elicit different avoidance tendencies. IA is not universal in academic reading but rather changes with the content of the article. Furthermore, it is difficult to describe students' reading attitudes and skills quantitatively. This study assessed students' IA quantitatively, focusing on students' reading habits rather than the content of the articles they read. This method has universality and is not affected by the article's content; hence, it is more suitable for studying IA in academic reading. Since students' reading habits are stable and do not change suddenly, it is possible to concretely explore and understand students' IA tendencies through data analysis. Based on this idea, eye-tracking technology can be used to describe the process of students' reading with more accurate data, which can be further analyzed to understand students' IA tendencies in the future.

Limitation

Due to the COVID-19 virus, the number of students participating in the the experiment was small, and individual differences may have had a greater impact on the experimental results.

In order to comply with research ethics regulations, we cannot combine students from two different classes and assign students with the same level of study to both the experimental and control groups. The results showed disparities in the pre-test scores between the experimental group and control group, which could potentially influence the interpretation of learning effects.

Abbreviation

IA Information Avoidance

Acknowledgements

I would like to thank Fuzheng Zhao for assistance with the e-book data download. I am grateful to the students who participated in the experiments.

Authors' contributions

JZ and SW drafted the initial manuscript and all the research. CY and LX provided insights into designing the experiment. CY provided supervision of the research.

Funding

A part of this research was supported by the Grants-in-Aid for Scientific Research Nos. [blinded for review] and [blinded for review] from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan Grant, file number 21H00905,22K13752.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 11 October 2023 Accepted: 13 November 2023 Published online: 27 November 2023

References

Chen, C.-M., & Kuo, C.-H. (2019). An optimized group formation scheme to promote collaborative problem-based learning. *Computers and Education, 133*, 94–115.

Covington, M. V., & Müeller, K. J. (2001). Intrinsic versus extrinsic motivation: An approach/avoidance reformulation. Educational Psychology Review, 13, 157–176.

- Ek, S., & Heinström, J. (2011). Monitoring or avoiding health information: The relation to inner inclination and health status. *Health Information and Libraries Journal*, 28(3), 200–209.
- Fuertes, M. C. M., Jose, B. M. D., Nem Singh, M. A. A., Rubio, P. E. P., & De Guzman, A. B. (2020). The moderating effects of information overload and academic procrastination on the information avoidance behavior among Filipino undergraduate thesis writers. *Journal of Librarianship and Information Science*, 52(3), 694–712.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1), 96–135.

Gwo-Jen Hwang, H.-C.C., & Yin, C. (2017). Objectives, methodologies and research issues of learning analytics. Interactive Learning Environments, 25(2), 143–146. https://doi.org/10.1080/10494820.2017.1287338

Hermida, D. et al. (2009). The importance of teaching academic reading skills in first-year university courses. SSRN 1419247.

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. Multimedia Tools and Applications, 80, 8091–8126.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). Oakland, CA, USA.

Maqtary, N., Mohsen, A., & Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning, 24*, 169–190.

- Mirjalili, S., & Mirjalili, S. (2019). Genetic algorithm. Evolutionary algorithms and neural networks: Theory and applications (pp. 43–55).
- Ren, Z., Uosaki, N., Kumamoto, E., Liu, G.-Z., & Yin, C. (2017). Improving teaching materials through digital book reading log. In *The 2017 international conference on advanced technologies enhancing education (ICAT2E 2017)* (pp. 90–96). Atlantis Press.
- Schilit, B. N., Golovchinsky, G., & Price, M. N. (1998). Beyond paper: Supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249–256).
- Soroya, S. H., Farooq, A., Mahmood, K., Isoaho, J., & Zara, S.-E. (2021). From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. *Information Processing and Management*, 58(2), 102440.
- Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. Review of General Psychology, 14(4), 340–353.
- Yin, C., Uosaki, N., Chu, H. C., Hwang, G.-J., Hwang, J., Hatono, I., & Tabata, Y. (2017). Learning behavioral pattern analysis based on students' logs in reading digital books. In *Proceedings of the 25th international conference on computers in education* (pp. 549–557).
- Zhou, J., & Yin, C. (2023). Information avoidance in educational technology. In 2023 international conference on artificial intelligence and education (ICAIE) (pp. 44–46). IEEE Computer Society.

Publisher's Note

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

Submit your manuscript to a SpringerOpen[™] journal and benefit from:

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

Submit your next manuscript at > springeropen.com