So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Australia: Asia-Pacific Society for Computers in Education

How Does The Quality of Students' Highlights Affect Their Learning Performance in e-Book Reading

Albert YANG[a*], Irene Y.L. CHEN[b], Brendan FLANAGAN[c] & Hiroaki OGATA[c]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Department of Accounting, National Changhua University of Education, Taiwan*
[c]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*yang.ming.35e@st.kyoto-u.ac.jp

Abstract: Measuring a student's highlight quality is a way to understand how well a student understands the learning materials. In general, students highlight the words or sentences they consider as important concepts during their learning process. An often way to measure students' highlight quality is to let graders give a highlight score after reading their highlights. However, when graders do not reach agreement on the quality of highlights and give different scores due to different background knowledge, the influence of highlight quality on learning performance may decrease. In this study, we explore the marker function on BookRoll and propose another way to grade students' highlight quality in terms of its similarity to the reference answer. Students can add highlights as markers when reading learning materials. The instructor will highlight the important concepts in the learning materials as a reference answer. Graders will use this reference answer to grade students' highlights. The average of students' quiz scores will be used to denote their learning performance. The results indicate that the highlight quality has a larger influence on students' learning performance than that of highlight frequency when graders evaluate the highlights along with a reference answer. Next, we divided students into two groups and performed a one-way ANOVA test to evaluate the influence of highlight quality on learning performance, and found that students who achieve high learning performance are also likely to achieve a high highlight score. We further compared the highlight quality and highlight frequency to see which indicator is more correlated with learning performance and found that highlight quality outperforms highlight frequency.

Keywords: e-Book reading, learning analytics, learning performance

1. Introduction

Using highlights is considered as a cognitive learning behavior which influences reading comprehension positively (Van Horne et al., 2016). Nian et al. (2019) and Yin et al. (2019) analyzed the frequency of e-reader functions on BookRoll such as NEXT, PREV and MARKER used by students to observe whether these behaviors are related to learning performance. They found that students who used highlights frequently had a tendency to achieve better learning performance. Al-khazraji (2019) let learners use highlights to mark the sentences they think are important, and found that the use of highlights greatly affects the effectiveness of learning. Yufan et al. (2020) proposed that in addition to analyzing the frequency of highlights, the area of highlights may also be related to learning performance. However, without considering the content of highlights, it is still possible to overuse or misuse them and decrease learning performance.

To overcome this issue, we conducted an experiment in a university course to investigate the influence of students' highlight quality on their learning performance.

RQ1: Is it better for graders to have a reference answer when grading students' highlights?
RQ2: What is the correlation between students' highlight quality and learning performance?
RQ3: Is highlight quality a better predictor than highlight frequency for predicting students' learning performance?

2. Experiment

2.1 Data Collection

In this study, we analyzed students' highlight records of a 12-week course in a university. A total of 22 slides of learning materials were uploaded to BookRoll. BookRoll is an e-Book reading system (Flanagan & Ogata, 2017) developed by Kyoto University. Students' e-Book reading actions in BookRoll have been introduced in detail by (Ogata et al., 2015) and (Flanagan & Ogata, 2018). There were 44 students enrolled in this course. Before each class, the instructor uploaded the slides of the class to BookRoll for students to preview. Students were asked to highlight the words or sentences they think are important using the marker function on BookRoll. There are two types of markers they can use, red marker for important concepts and yellow marker for concepts they feel difficult to understand. The instructor gave two graders a reference answer. Graders graded the highlight to see whether students highlight the important concepts. Figure 1 is an example of marker function on BookRoll. Finally, Students were given a quiz every week to measure their understanding level. The average of quiz scores were used to denote their learning performance.

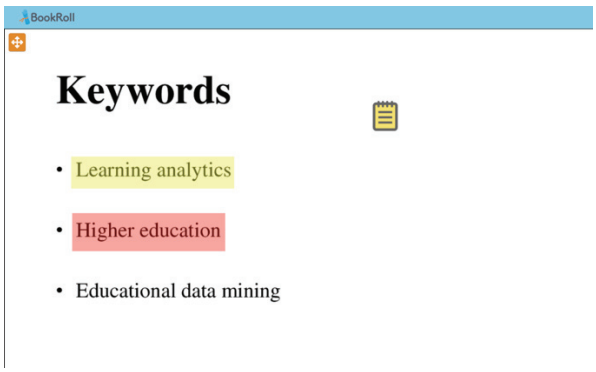

Figure 1. An example of marker function on BookRoll.

2.2 Grading Highlight Score with Reference Answer

Two graders independently graded students' highlights. They first graded the highlights based on their background knowledge and understanding about the course. Then they used the highlights from the instructor as a reference answer and graded the highlights again. To know the level of agreement on the highlight scores from different graders, the Cohen-Kappa-Coefficient of their grading result were being calculated.

2.3 Highlight Quality vs. Highlight Frequency

Frequency of adding highlights is often discussed when exploring the correlation between highlight function and students' learning performance. In this study, highlight quality is measured by the average of highlight scores grade by humans, whereas highlight frequency is denoted as the number of highlights students add. Finally, learning performance is measured by the average of quiz scores. We compared highlight quality with highlight frequency to decide which indicator has a better correlation with learning performance.

2.4 Procedure

To answer our three research questions, we used Cohen-Kappa-Coefficient to compare two different grading approaches and evaluated which of it better reflects students' learning performance. After that, a one-way ANOVA test is conducted to observe the impact of highlight quality on learning

performance. Finally, we measured whether highlight quality is a better indicator than highlight frequency for predicting learning performance. The result and discussion for each research question is listed as the followings.

3. Results and Discussion

For Research question 1, Is it better for graders to have a reference answer when grading students' highlights?

Table 1 and Table 2 show the result of highlight scores graded by two graders with a reference answer and without a reference answer. The diagonal indicates that two graders gave the same score. The sum of diagonal scores (25) in Table 1 is larger than the sum of diagonal scores (17) in Table 2, which means the grading result is more consistent between two graders when a reference answer is given. Table 3 shows the Cohen-Kappa-Coefficient and the mean absolute error of two highlight scores, with reference answer and without reference answer, and the Spearman correlation between two grading approaches and students' learning performance. When graders grade with their own knowledge level, the agreement level is only 0.179, which means the possibility of the agreement is probably occurring by chance. On the other hand, the agreement level of highlight score graded with a reference answer is satisfactory with 0.423, which indicates that the possibility of the agreement does not occur by chance. The mean absolute error (MAE) between two graders was calculated for both approaches to measure the average difference between graders. The MAE between highlight scores is 6.61 when a reference answer is provided and 10.84 when a reference answer is not provided. The result indicates that the difference between highlight scores is larger on average when a reference answer is not provided. The Spearman correlation between learning performance and highlight score with a reference answer is 0.779, which is also better than the result of 0.525 without a reference answer. From this result, we know that when graders follow a reference instead of relying on their own background knowledge to grade, the highlight score is more consistent and correlates better to the learning performance. Table 4 lists students' highlight score, highlight frequency and their learning performance. The highlight score is computed by the average of two graders' scores based on the reference answer.

Table 1. *The result of highlight score based on reference answer.*

		Grader2			
	Highlight Score	>=90	>=80	>=70	<70
Grader1	>=90	11	2	2	0
	>=80	0	4	4	1
	>=70	0	1	4	6
	<70	0	1	2	6

Table 2. *The result of highlight score without reference answer.*

		Grader2			
	Highlight Score	>=90	>=80	>=70	<70
Grader1	>=90	8	3	1	3
	>=80	2	1	2	4
	>=70	1	4	4	2
	<70	1	1	3	4

Table 3. *The Cohen-Kappa-Coefficient for two grading approaches and the Spearman correlation between two grading approaches and the learning performance.*

	With Reference Answer	Without Reference Answer
Cohen-Kappa-Coefficient	0.423	0.179
Mean Absolute Error	6.613	10.84
Spearman Correlation	0.779***	0.525***

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

Table 4. *Students' highlight score, highlight frequency and learning performance.*

Student	HS	F	P	Student	HS	F	P
S1	81.5	524	84	S23	72.5	376	94
S2	60	333	64	S24	65	333	75
S3	60	434	66	S25	73	439	72
S4	67	392	77	S26	71	328	70
S5	67	381	74	S27	76	323	60
S6	67	498	73	S28	62	344	78
S7	84.5	466	81	S29	88	365	86
S8	90	339	81	S30	90	323	89
S9	89.5	360	89	S31	88	333	92
S10	76.5	349	80	S32	85	318	84
S11	63.5	360	70	S33	98	397	95
S12	60	318	62	S34	99	333	90
S13	60	323	60	S35	87	530	97
S14	86	471	88	S36	79.5	318	77
S15	89.5	392	82	S37	65	318	81
S16	97	530	100	S38	78.5	344	83
S17	93.5	402	86	S39	86	434	78
S18	91	439	95	S40	63.5	355	77
S19	100	450	97	S41	93.5	344	85
S20	79	376	80	S42	89	386	100
S21	91	450	90	S43	78	530	65
S22	82.5	434	100	S44	74	318	60

HS=Highlight Score, F=Frequency, P=Learning Performance

For Research question 2, What is the correlation between students' highlight quality and learning performance?

After deciding the approach to grade highlights, we want to know the influence of highlight quality on learning performance. We classified the students into two groups based on their learning performance. Students with top 20% learning performance were assigned to HIGH_PERFORMANCE, and the rest of students were assigned to LOW_PERFORMANCE. Figure 2 is a boxplot of the highlight score for two groups. It demonstrates that most students in HIGH_PERFORMANCE have a better maker score than students in LOW_PERFORMANCE. Since both groups represent the normal distribution, a one-way ANOVA test is conducted to observe whether highlight score influences students' learning performance. Table 5 indicates that the mean of highlight score in HIGH_PERFORMANCE is larger than LOW_PERFORMANCE at a statistical significance level. The result shows that students with high learning performance are likely to achieve a better highlight score. The plausible explanation is that they know what they know and what they do not know, and they are capable of highlighting the sentences or words that represent the important concepts in learning

materials.

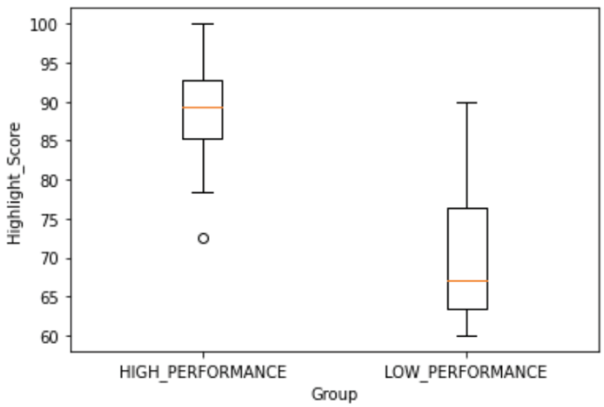

Figure 2. A boxplot of the highlight score for two groups.

Table 5. *The one-way ANOVA test for group with high highlight quality and group with low highlight quality.*

	Mean		*SD*		*F*
	HP	*LP*	*HP*	*LP*	
HIGHLIGHT_SCORE	*88.86*	*70.15*	*6.59*	*8.56*	*62.89****

***: p < 0.001, **: p < 0.01, *: p < 0.05; HP=High Performance, LP=Low Performance

For Research question 3, Is highlight quality a better predictor than highlight frequency for predicting students' learning performance?

Next, we want to know the correlation between the frequency of adding highlights and the highlight score, and whether the quality of highlights is more related to students' learning performance than the frequency of adding highlights. Table 6 shows the Spearman correlation between highlight score, frequency of adding highlights and learning performance. The correlation between highlight score and highlight frequency is only 0.31, which means students who frequently add highlights may not achieve a high highlight score if they missed the important concepts or marked the concepts that are considered as less important. The possible reason is that they are not sure about what they know or what they do not know, which leads to the misuse of highlights. The result also shows that the score of highlights is more correlated with learning performance than that of frequency. This suggests that students who are able to grasp the important concepts in learning materials are more likely to achieve higher learning performance.

Table 6. *The Spearman correlation between highlight score, highlight frequency and students' learning performance.*

	HS	HF	LP
HS	1		
HF	0.31*	1	
LP	0.77***	0.37*	1

***: p < 0.001, **: p < 0.01, *: p < 0.05; HS=Highlight Score, HF=Highlight Frequency, LP=Learning Performance

4. Conclusion

In this research, we conducted an experiment on two highlight grading approaches to investigate whether grading based on a reference answer is better than grading based on graders' own knowledge level. The result shows that agreement on score is hard to achieve for graders without a reference answer. After finding the better grading approach, we performed a one-way ANOVA test to examine whether highlight quality has influence on students' learning performance. The results demonstrated

that students who achieve high learning performance are more likely to achieve a high highlight score because they can highlight the important concepts and identify concepts they are not familiar with. Conversely, learners who are not able to get high learning performance also cannot achieve a high highlight score. They are likely to mark the concepts that are not important since they are not sure what they know and what they do not know. Finally, we measure the correlation among highlight quality, highlight frequency and learning performance. We calculated the Spearman correlation for three descriptive data. The correlation between highlight quality and highlight frequency is low, which means students who mark frequently are not guaranteed to achieve a high highlight score because they are not able to identify important concepts from the learning materials and misuse the highlight function. Also, the correlation between learning performance and highlight quality is higher than the correlation between learning performance and highlight frequency, which indicates that highlight quality is a better predictor than highlight frequency for predicting learning performance.

In future work, we want to explore more methods which leverage machine learning algorithms for grading highlights. For instance, a semi-automatic way in which the instructor provides a reference answer and students' highlights are automatically graded using machine learning algorithms, or a full-automatic approach that applies machine learning algorithms to summarize the text as important concepts and the highlights can be automatically graded.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Van Horne, S., Russell, J. E., & Schuh, K. L. (2016). The adoption of mark-up tools in an interactive e- textbook reader. Educational Technology Research and Development, 64(3), 407-433.

Yin, C.,Yamada, M., Oi,M., Shimada, A.,Okubo, F.,Kojima, K.,Ogata,H. (2019). Exploring the Relationships between Reading Behavior Patterns and Learning Outcomes based on Log Data from e-books: A Human Factor Approach, International Journal of Human-Computer Interaction, 313-322.

Al-khazraji, A. (2019). Analysis of Discourse Markers in Essays Writing in ESL Classroom. International Journal of Instruction, 12(2), 559-572.

Yufan, X., Xuewang, G., Li, C., Satomi, H., Yuta, T., Hiroaki, O., & Yamada, M. Can the Area marked in eBook Readers Specify Learning Performance. Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20) (pp. 638-648)

Nian, M. W., Lee, Y. H., & Wu, J. Y. (2019). Using machine learning to explore the associations among e-reader operations and their predictive validity of learning performance. International Conference on Learning Analytics & Knowledge (LAK19) (pp. 1-6).

Flanagan, B., & Ogata, H. (2017, November). Integration of learning analytics research and production systems while protecting privacy. In The 25th International Conference on Computers in Education, Christchurch, New Zealand (pp. 333-338).

Flanagan, B., & Ogata, H. (2018). Learning analytics infrastructure for seamless learning. Companion Proceedings 8th International Conference on Learning Analytics & Knowledge (LAK18) (2018), http://hdl.handle.net/2433/233071

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015, January). E-Book-based learning analytics in university education. In International conference on computer in education (ICCE 2015) (pp. 401-406).