Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Investigation on Practical Effects of the Explanation in a K-12 Math Recommender System

Yiling DAI[a*], Kyosuke TAKAMI[a], Brendan FLANAGAN[a] & Hiroaki OGATA[a]
[a]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*dai.yiling.4t@kyoto-u.ac.jp

Abstract: Providing explanations in educational recommender systems is expected to improve students' motivation and learning outcomes. In this study, we aim at exploring the practical effects of the explanation in a K-12 math recommender system. We implemented a concept-explicit quiz recommender system and generated explanations of the estimated quiz difficulty and the expected learning gains on the concepts. We then conducted an A/B experiment (N=281) in a Japanese high school to verify the explanation's effects on students' acceptance of the recommendations and related learning behaviors. We found that the explanations led to a higher click rate of recommendations. Besides, among the group displayed with explanations, the students who used the system more frequently and consistently gave more positive feedback on the explanations.

Keywords: Recommender system, math quiz, explanation, recommendation acceptance, A/B experiment

1. Introduction

Recommending learning materials based on individual student's knowledge state has become an essential task in learning management systems. Unlike consuming entertaining products such as movies or music, reading learning materials or solving quizzes requires higher level of motivation and determination of users. Therefore, it is important to provide explanations to the students when recommending learning materials. Some pioneering works have explored the effects of explanations in educational recommender systems and found the explanations may: increase the students' attention towards recommended practices, the willingness to open them (Barria-Pineda et al., 2021); increase the students' trust in hints, perceived usefulness of them, and the intention to use them again (Conati, Barral, Putnam, & Rieger, 2021); increase the students' perceived unexpectedness, novelty of recommended courses, and the interests in them (Yu, Pardos, Chau, & Brusilovsky, 2021); increase the students' clicks of the recommended quizzes (Takami, Dai, Flanagan, & Ogata, 2022). However, it is still under-researched how the explanations affect students' learning behaviors and performance in real classroom settings especially at the level of K-12 education. In this study, we aim at enriching this research topic by exploring the real-world effects of explanations in a K-12 math quiz recommender system.

In the context of learning math, Birenbaum et.al. (1993) suggested that identifying specific misconcepts and difficult areas is more instructive than a test score for remediation. Therefore, it is important to recommend math quizzes that address the students' weak points which are readable math concepts. Previously, we proposed a simple and explainable math quiz recommender system in which estimations are made based on human-readable concepts (Dai, Flanagan, Takami, & Ogata, 2022). We recommended quizzes that are appropriately difficult and help address the student's weak areas, companioned with the corresponding explanations. The explanation is expected to increase the students' belief of their ability to solve the quiz and the perceived values of solving the quiz, which are suggested as strong predictors of learning motivation (Wigfield & Eccles, 2000). On the other hand, the explanations on the math concepts are expected to help the students construct relationships between the quizzes they are going to solve and the math knowledge they already possess, which is consistent with the subsumption theory proposed by Ausubel (1962).

To verify the learning effects of the explanations, we conducted an A/B experiment in a Japanese high school. We focused on the following research questions:
RQ1: Do the explanations motivate the students to accept the recommendations?
RQ2: Do the students who favor the explanations have any characteristics in learning behavior?

The remaining of this paper is structured as: In Section 2, we introduce previous works about explainable recommender systems in educational context. Section 3 overviews the design and implementation of the concept-explicit math quiz recommender system. In Section 4, we report the real classroom experiment and discuss the findings of the explanation's effects. We conclude this study in Section 5.

2. Related Work

In general recommender systems, providing explanations on why the item is recommended is supposed to raise the user's trust (X. Wang et al., 2018), promote the user to accept the recommendation (Huang et al., 2019; X. Wang et al., 2018; Xie, Hu, Cai, Zhang, & Chen, 2021), and support the user in making a good decision (N. Wang, Wang, Jia, & Yin, 2018; X. Wang et al., 2018). In educational recommender systems, many researchers have attempted to generate explanations for a recommended learning material (Barria-Pineda et al., 2021; Conati et al., 2021; Rahdari et al., 2020; Takami et al., 2022; Yu et al., 2021). However, many works did not include an experiment evaluating the effects of the explanations, such as (Rahdari et al., 2020) in the educational context. Some of the works only conducted simulation-based experiments (Bilgic & Mooney, 2005; Conati et al., 2021; N. Wang et al., 2018; X. Wang et al., 2018). For example, Conati et.al. (2021) conducted a control experiment to investigate participants' feedback toward the explanations in a laboratory environment which lasted for two or three hours. We consider that it is insufficient to conclude something about the long-term effect in learning. Yu et.al. (2021) compared the effects of three types of explanations in a course recommender system. We consider that course selection is different from the scenario of doing math quizzes as the latter one requires users' continuous engagement with the system. To the best of our knowledge, only two works evaluated the explanation's practical effect in real settings: Barria-Pineda et.al. (2021) conducted a long-term experiment to test the learning effects of explanations on recommending programming exercises to university students. Different from their experiment setting where both groups were exposed to the explanation, our study verifies the effect of explanation by comparing groups with and without explanations. Takami et.al. (2022) explored the explanation's effect on the usage of recommended math quizzes. Our study focuses on a different recommender system and succeeded to collect more log data and analyze in more depth.

3. Overview of Our Concept-Explicit Quiz Recommender System

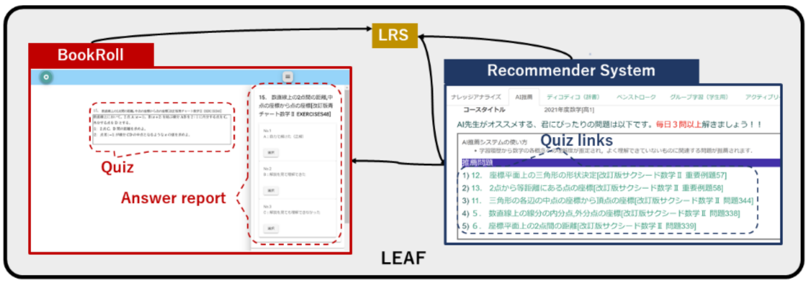

Figure 1. The overview of the system.

In our previous work (Dai et al., 2022), we proposed a concept-explicit quiz recommender system where the estimations are made based on the math concepts. We recommend math quizzes and generate explanations based on two strategies: 1) quizzes of appropriate difficulty. 2) quizzes that improve students' weak points. We implemented the recommender system under the Learning Evidence Analytics Framework (LEAF) (Flanagan & Ogata, 2018) where an ebook reading module BookRoll is used to display the math quizzes and collect the students' answers, and a learning record storage module

LRS is used to record the learning activities in the system. As illustrated in Figure 1, our recommender system guides the students to the BookRoll page of the quiz based on the students' log data generated in the system.

4. Practical Effects in Real Classrooms

4.1 Experiment Design

To answer the research questions, we conducted a quasi-experiment (Cohen, Manion, & Morrison, 2011) where the experimental group was exposed to the recommender system with explanations and the control group was exposed to the recommender system without explanations. As Figure 2 shows, the treatment of the control group only displays five links of the recommended quizzes. In addition to the links, the treatment of the experimental group contains i) a concept panel that demonstrates the student's current mastery level, ii) the estimated difficulty level of the quiz, and iii) the expected mastery level update. Note that this is one of various possible visualizations and exploring the effects of different visualizations is out of the scope of this study.

Figure 2. The treatment of the control group and experimental group.

A Japanese high school adopts the LEAF system as a supplement of its brick and mortar education. 7 classes of first-year students (N=281) were invited to participate in the experiment and we obtained their consent on providing the learning data for research use. The 7 classes were separated into control (4 classes) and experimental (3 classes) groups randomly. During the experiment period (a short school holiday from February 9, 2022 to February 20, 2022), the students were required to complete at least 20 assigned quizzes and 10 recommended quizzes for reviewing a previously learnt topic. The target topic "geometry and equations" contains 262 descriptive problems from two quiz books. Normally, the students are supposed to complete the quizzes at their own paces and the teachers give supervision or intervention when they think necessary. Thus, it is always the case that an average student cannot finish all the quizzes and there is a need to personalize the quizzes of a doable size. Note that the students have both the paper-version and digital-version of the quizzes at hand, and it is observed they are more comfortable with the traditional paper version.

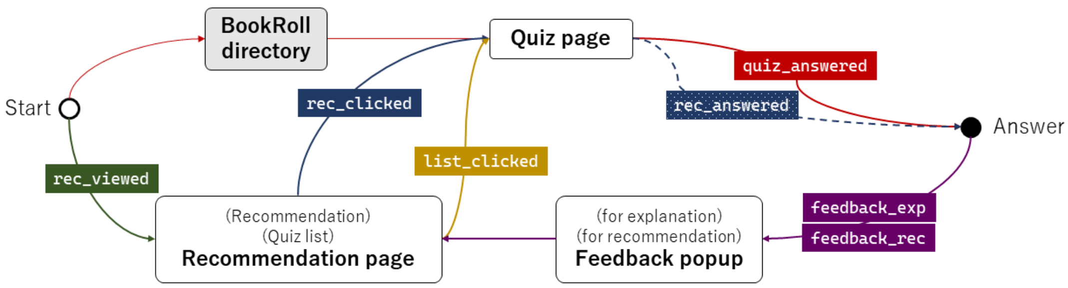

Figure 3. The student workflow and the collected log data.

Figure 3 illustrates the student workflow and the collected log data in our system. To access a quiz page, the student can directly open one topic from the directory and move around the pages in the topic. If the student chooses to answer the recommended quizzes, s/he can open the recommendation page, where a recommendation block displays the recommendations and associated explanations, and a quiz list block displays all the links to the quizzes in the current topic. Every time a student opens the recommendation page, some recommendations are generated and recorded as *rec_viewed*. The clicks of the links in the recommendation block and quiz list block are recorded as *rec_clicked* and *list_clicked*, respectively. After a quiz page is opened, the student is supposed to solve the quiz either using the digital pen or their own notes, check the answer, and then report whether the answer is correct or not. Only 0 (incorrect) or 1 (correct) is recorded as *quiz_answered*. If the *quiz_answered* action happens after a *rec_clicked* in a predefined time window (one hour in this study), the *quiz_answered* is also counted as an answered recommendation *rec_answered*. After the student answers a quiz from recommendation, s/he is required to rate the explanation (if displayed) and the recommendation with a score from 1 to 5, which is recorded as *feedback_exp* and *feedback_rec*, respectively.

4.2 Data Collection and Preprocessing

We used the following process to preprocess the log data:

1. For each student, we sorted her/his log data in temporal sequence.
2. We treated an action as a start of a new session if it happened more than one hour later than its previous action.
3. We computed the average time spent on one quiz in each session. We then deemed a session as valid if the time spent on one quiz is longer than one minute.
4. We excluded all the actions in invalid sessions.

After the filtering process, we had 30,673 records of 184 unique students and 341 unique quizzes remained in our dataset. We processed the data mainly using Python and jamovi.

4.3 Results and Discussion

4.3.1 Explanation's Effects Over Students Who Used the Recommendations

Table 1. *Welch's T-test of Students' Recommendation Usage in Experimental and Control Groups*

Variable (per student)	Group	N	Mean	SD	Welch's T-test Statistics	df
Recommendation click rate	**Control**	**71**	**0.154**	**0.178**	**2.511****	**90.7**
	Experimental	**58**	**0.264**	**0.290**		
Recommendation answer rate	Control	70	0.817	0.156	1.382	121.8
	Experimental	55	0.853	0.135		
Perception of recommendation	Control	69	3.592	0.896	0.662	123.7
	Experimental	57	3.690	0.779		

Hypothesis: Control<Experimental
* p<.05, ** p<.01, *** p<.001

To answer **RQ1**, we are interested in knowing whether the explanation motivates the students to accept the recommended quizzes and perceive the recommendations positively. To do so, we first computed recommendation click rate as $\frac{\#\ \text{rec\_clicked}}{\#\ \text{rec\_viewed}}$, and recommendation answer rate as $\frac{\#\ \text{rec\_answered}}{\#\ \text{rec\_clicked}}$, respectively. We also took the average of the *feedback_rec* as a student's overall perception of the recommendations. We then conducted a Welch's T-test on the values between two groups. As we can see from Table 1, there is a significant difference between the experimental (0.264) and control (0.154) groups in terms of the recommendation click rate. This indicates that the student is more likely to click a recommendation if the explanation is provided, which provides a positive answer to **RQ1**. Our result is a reinforcement of the findings in Takami et.al.'s work (2022) as we analyzed the click rate at the level of every individual student, which excludes the biases brought by specific "heavy" users.

The answer rate was higher than 0.8 in both groups. In other words, if the student clicks the quiz, s/he always answers the quiz given this learning context. To this end, raising the students' attention and interest toward a quiz at an earlier stage is more important for increasing the recommendation acceptance rate. In terms of the perception of the recommendation, we did not find a significance between two groups. The reason could be complicated as the goodness of the recommendation is difficult to define at the first place. Unlike watching a movie, doing a math quiz is not always enjoyable. The students' feedback on the recommendation may be influenced by their values and attitudes towards math learning. In the future work, we need to collect more specified students' feedback on the recommendation.

4.3.2 Explanation's Relationship with Learning Characteristics

Table 2. *Descriptive Statistics and Correlations for Perception of the Explanation and Learning Behavior Characteristics*

		Descriptive statistics			Correlations
	Variable	N	Mean	SD	**1**
Perception of explanation	1. Mean of *feedback_exp*	57	3.72	0.797	**-**
Characteristic of learning behavior	2. *mean_quiz_time* (minute)	184	9.59	6.52	**0.286* (N=57)**
	3. *num_days*	184	2.52	1.66	**0.263* (N=57)**
	4. *max_cons_days*	184	1.41	0.733	**0.431*** (N=57)**

* p<.05, ** p<.01, *** p<.001
Correlations among Variables 2-4 are omitted since they are not our focus.

Among the students who were displayed with the explanations, who favors the explanations from the perspective of learning behavior? Utilizing the session information, we obtained three metrics which indicate different characteristics of the learning behavior: a) average time spent on answering a quiz, b) number of days using the system, and c) maximum number of days using the system consecutively. We took the average of the *feedback_exp* as the student's overall perception of the explanation. Table 2 shows the descriptive statistics and the correlations of the four variables. We found that all of the three metrics are positively related to the perception of the explanation. Especially, the students who used the system more consistently (i.e., larger number of consecutive days) tended to have a better perception of the explanation. To this end, we answer **RQ2** as the students who favored the explanations used the system more consistently. Although the causality remains unclear, we think this finding shows evidence for Bilgic and Mooney's argument (2005)--- explanations could improve user's satisfaction, foster user's trust and thus motivate user's continuous use of the system. In the future, we need to further investigate how the students' personalities and other learning strategies are related to their perceptions of the explanations and their usage of the recommendation.

5. Conclusions and Future Work

In this study, we focused on investigating the practical effects of a K-12 math quiz recommender system where explanations are generated to help students better understand why the quiz is recommended to them. We implemented the recommender system in an integrated learning framework and generated explanations of the estimated difficulty of the quiz and the expected master level update on the student's weak concepts. We then conducted a 12-day A/B experiment (N=281) in a Japanese high school to verify the explanation's effects on students' acceptance of recommendations and related learning behaviors. We found promising results such as: a significant difference between the click rates in groups with and without explanations; students who highly valued the explanations spent much time on each quiz, used the system frequently and consistently.

Some limitations need to be addressed in future works. In this study, we mainly investigated the explanation's effect on the acceptance of the recommendation. We need to further explore the effects on learning performance such as whether the weak points are addressed or not. Besides, we need to analyze in what sequences the quizzes were recommended to the students and how did their mastery levels evolve during the usage of the recommender system. In addition, to personalize the recommendation mechanism and the corresponding explanations, we need to collect and analyze the data that indicate the students' personalities, emotions, attitudes towards learning, learning strategies, and so on.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722, JSPS Grant-in-Aid for Scientific Research (Exploratory) 21K19824, NEDO JPNP20006 and JPNP18013.

References

Ausubel, D. P. (1962). A Subsumption Theory of Meaningful Verbal Learning and Retention. *The Journal of General Psychology*, *66*(2), 213–224.

Barria-Pineda, J., Akhuseyinoglu, K., Želem-Ćelap, S., Brusilovsky, P., Milicevic, A. K., & Ivanovic, M. (2021). Explainable Recommendations in a Personalized Programming Practice System. *Proceedings of International Conference on Artificial Intelligence in Education*, 64–76.

Bilgic, M., & Mooney, R. J. (2005). Explaining Recommendations: Satisfaction vs. Promotion. *Proceedings of Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research at the 2005 International Conference on Intelligent User Interfaces*.

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. K. (1993). Diagnosing Knowledge States in Algebra Using the Rule-Space Model. *Journal for Research in Mathematics Education*, *24*(5), 442–459.

Cohen, L., Manion, L., & Morrison, K. (2011). Experiments, Quasi-Experiments, Single-Case Research and Internet-based Experiments. In *Research Methods in Education (7th ed.)* (pp. 312–334).

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, *298*, 103503.

Dai, Y., Flanagan, B., Takami, K., & Ogata, H. (2022). Design of a User-Interpretable Math Quiz Recommender System for Japanese High School Students. *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge.*

Flanagan, B., & Ogata, H. (2018). Learning Analytics Platform in Higher Education in Japan. *Knowledge Management & E-Learning*, *10*(4), 469–484.

Huang, X., Fang, Q., Qian, S., Sang, J., Li, Y., & Xu, C. (2019). Explainable Interaction-Driven User Modeling over Knowledge Graph for Sequential Recommendation. *Proceedings of the 27th ACM International Conference on Multimedia (MM'19)*, 548–556.

Rahdari, B., Brusilovsky, P., Thaker, K., & Barria-Pineda, J. (2020). Using Knowledge Graph for Explainable Recommendation of External Content in Electronic Textbooks. *Proceedings of the Second International Workshop on Intelligent Textbooks 2020 Co-Located with 21st International Conference on Artificial Intelligence in Education (AIED 2020)*, *2674*, 50–61.

Takami, K., Dai, Y., Flanagan, B., & Ogata, H. (2022). Educational Explainable Recommender Usage and Its Effectiveness in High School Summer Vacation Assignment. *Proceedings of the 12th International Learning Analytics and Knowledge Conference (LAK22)*, 458–464.

Wang, N., Wang, H., Jia, Y., & Yin, Y. (2018). Explainable Recommendation via Multi-Task Learning in Opinionated Text Data. *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'18)*, 165–174.

Wang, X., Chen, Y., Yang, J., Wu, L., Wu, Z., & Xie, X. (2018). A Reinforcement Learning Framework for Explainable Recommendation. *Proceedings of 2018 IEEE International Conference on Data Mining (ICDM)*, 587–596.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Xie, L., Hu, Z., Cai, X., Zhang, W., & Chen, J. (2021). Explainable recommendation based on knowledge graph and multi-objective optimization. *Complex & Intelligent Systems*, *7*(3), 1241–1252.

Yu, R., Pardos, Z., Chau, H., & Brusilovsky, P. (2021). Orienting Students to Course Recommendations Using Three Types of Explanation. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 238–245.

Zhang, Y., & Chen, X. (2020). Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval*, *14*(1), 1–101.