

RESEARCH

Free and Open Access

Beyond recommendation acceptance: explanation's learning effects in a math recommender system

Yiling Dai ^{1*}, Kyosuke Takami ², Brendan Flanagan ³ and Hiroaki Ogata ¹

*Correspondence:
dai.yiling.4t@kyoto-u.ac.jp
Academic Center for Computing
and Media Studies,
Kyoto University, Kyoto, Japan
Full list of author information is
available at the end of the article

Abstract

Recommender systems can provide personalized advice on learning for individual students. Providing explanations of those recommendations are expected to increase the transparency and persuasiveness of the system, thus improve students' adoption of the recommendation. Little research has explored the explanations' practical effects on learning performance except for the acceptance of recommended learning activities. The recommendation explanations can improve the learning performance if the explanations are designed to contribute to relevant learning skills. This study conducted a comparative experiment (N = 276) in high school classrooms, aiming to investigate whether the use of an explainable math recommender system improves students' learning performance. We found that the presence of the explanations had positive effects on students' learning improvement and perceptions of the systems, but not the number of solved quizzes during the learning task. These results imply the possibility that the recommendation explanations may affect students' meta-cognitive skills and their perceptions, which further contribute to students' learning improvement. When separating the students based on their prior math abilities, we found a significant correlation between the number of viewed recommendations and the final learning improvement for the students with lower math abilities. This indicates that the students with lower math abilities may benefit from reading their learning progress indicated in the explanations. For students with higher math abilities, their learning improvement was more related to the behavior to select and solve recommended quizzes, which indicates a necessity of more sophisticated and interactive recommender system.

Keywords: Recommender system, Explainable recommender system, Educational recommender system, Learning performance, Math learning



© 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

In a traditional school environment, students get advice on what content to review, what exercise to practice from their teachers, peers, or parents. This advice can be viewed as recommendations and is easy to accept as it is provided from trustworthy people around the student. However, this type of recommendation is not always available as these people are not full-time personal advisors for an individual student. With the adoption of information technologies in education, it is possible for a system to make personalized recommendations of learning activities and materials through complex computations based on the learning data it collects (da Silva et al., 2023). In a broader scope, such recommender systems have been criticized as being black boxes in terms of how they get to the decision that the recommendation is necessary for the user (Shin, 2021). Providing explanations is an effective way to increase the transparency, persuasiveness, effectiveness, trustworthiness, and satisfaction of the recommender system (Zhang & Chen, 2020).

In terms of being explainable, educational recommender systems have some distinct needs, such as meeting developmental needs of learners (G. Wang et al., 2022), serving various stakeholders including teachers and parents (Murgia et al., 2019), and supporting learners' metacognitive processes of controlling, monitoring, and planning (Khosravi et al., 2022). Some researchers have attempted to generate explanations for the recommendations (Barria-Pineda et al., 2021; Conati et al., 2021; Rahdari et al., 2020; Takami et al., 2022; Yu et al., 2021), aiming to improve the acceptance of the recommendations, the perceptions of the system, and learning performance. As da Silva et al. (2023) pointed out, the online evaluation was under-researched in the area of educational recommender systems due to the complexity and high cost of conducting a real-life experiment. To the best of our knowledge, only few research (Barria-Pineda et al., 2021) had evaluated the effectiveness beyond the acceptance of the recommendations. As reading more recommended materials or solving more recommended quizzes does not necessarily relate to a higher learning performance, it is necessary to explore how the explanations can improve learning performance beyond the acceptance of the recommendation.

In this study, we focus on the scenario of learning math by practicing quizzes that are problems requiring the students to reach the answer in a step-by-step manner. When having a large repository of quizzes to practice, a recommender system can help detect the students' knowledge states and identify appropriate quizzes to address a students' weak points efficiently. Providing explanations of such recommender systems can not only improve the transparency of the model, but also improve the learning performance if the explanations are designed to contribute to relevant learning skills. We selected the concept-explicit explainable recommender system (Dai, Flanagan, et al., 2022) as our target model for the following reasons: 1) the model recommends math quizzes that are supposed to address students' weak concepts, which is considered important in learning math (Birenbaum et al.,

1993). 2) The recommender system estimates students' current mastery levels of math concepts and provides visual and textual explanations on this. In learning, knowing what one knows and what one does not know, which is also named as meta-cognitive skill, plays an important role in cognitive learning outcomes (Bahri & Corebima, 2015; Veenman et al., 2014). 3) The recommender system recommends quizzes that contribute to the students' current learning states and provides visual and textual explanations on this. We suppose the explanations help the students understand the connections between the newly encountered knowledge and the already known knowledge, which is considered an important process in learning (Ausubel, 1962).

We then investigate the effect of the concept-explicit explainable recommender system on learning performance, perceptions of the system, and quiz solving behaviors in the system. Students' perceptions of the system and quiz solving behaviors were also examined as they have the potential to provide insight into the aspects of the recommender that were influential in outcomes achieved by using the system. Consequently, we address the following research questions:

RQ1: Do recommendation explanations improve students' learning performance in terms of the summative assessment of mathematical concepts?

RQ2: Do recommendation explanations improve students' perceptions of the recommender system and attitudes towards math learning?

RQ3: Do recommendation explanations encourage the students to solve more quizzes?

RQ4: What is the relation of explanation-related learning activities (e.g., viewing the explanations, clicking the recommendations) in the recommender system and learning improvement, and is the relation different between high and low math ability groups?

Related work

Explainable recommender systems and their effects

Recommender systems make recommendations from a large set of available items based on users' preferences (Bobadilla et al., 2013). With the fast development of information technology, such recommender systems have become pervasive in various domains such as government, business, and education (Lu et al., 2015). However, recommender systems have been criticized as being black boxes in terms of how they decide a user may like an item (Shin, 2021). Research efforts have been observed in developing explainable recommender systems that not only provide users with recommendations on what information to consume, but also the reason why the recommendation is considered to be relevant to users' interests (Zhang & Chen, 2020). Providing explanations is supposed to increase users' trust towards the system (Huang et al., 2019; Kizilcec, 2016), acceptance

of recommendations (Cramer et al., 2008; Huang et al., 2019; Xie et al., 2021), and help users make an informed decision (Bilgic & Mooney, 2005; N. Wang et al., 2018; X. Wang et al., 2018; Xian et al., 2019).

Compared with the explainable recommender systems applied in a broader scope, educational recommender systems have some distinct needs related to teaching and learning. Wang et al. (2022) pointed out that meeting developmental needs of learners is an important principle when designing an artificial intelligent system, which suggests providing the learners with training opportunities to articulate the system. Murgia et al. (2019) argued that the explanations of educational recommender systems should serve multiple stakeholders including learners, teachers, and parents. Khosravi et al. (2022) considered this problem related to learning science as the explanations may improve learners' metacognitive processes such as controlling, monitoring, and planning. In this study, we stand in the same line as we consider the explanations not only increase the chances of students to practice the recommended quizzes, but more importantly, help them track their learning progress, select the appropriate recommendations, and achieve good learning performance.

Evaluation of the effects of explainable recommender systems

Following Erdt et al.'s (2015) classification, there are mainly three goals for evaluating educational recommender systems:

- recommender system performance, focusing on the accuracy of the recommendations, which is usually measured by a large existing dataset, namely, offline evaluation;
- user-centric elements, focusing on how learners perceive the system and whether they are satisfied with the system, which can be measured by user studies such as surveys and interviews;
- and learning performance, emphasizing on the ultimate goal of adopting the recommender system, namely, whether the learning achievement or the learning efficiency is improved. This is measured by real-life experiments which require a longer span of evaluation time and more system support.

According to Erdt et al.'s (2015) literature review, only around 10% of works conducted real-life evaluations during 2000 to 2014. This trend can also be observed if we further narrow down to the evaluation of the explanations in educational recommender systems. In Table 1, we summarize the recent works on explainable recommender systems who conducted real-life experiments to evaluate the system. As highlighted in bold letters, only a number of limited studies explored the system's effects on learning performance and conducted real-life experiments more than one time. Note that Barria-Pineda et al.'s work (2021) conducted a semester-long real-life experiment and found that the students invested their time to read the explanations. However, their results cannot directly answer whether

Table 1 Evaluations of the effects of educational explainable recommender systems

Previous work	Evaluation metric	Evaluation method
Karga & Satratzemi (2019)	Acceptance of the recommendation, perception of decision effectiveness, decision confidence, etc.	One-time online real-life experiment
Conati et al. (2021)	Usage of the explanation, perception of the explanation, learning gain	One-time laboratory real-life experiment
Yu et al. (2021)	Perception of the novelty, unexpectedness, and the acceptance of the recommendation	One-time laboratory real-life experiment
Barria-Pineda et al. (2021)	Learning activity engagement, learning performance	Semester-long real-life experiment
Takami et al. (2022), Dai, Takami, et al. (2022)	Acceptance of the recommendation	Week-long real-life experiment

the explanations help improve learning performance for two reasons: 1) Their experiment did not include a control/experimental setting to conduct direct comparisons. 2) They adopted regression models to explore the relationships between average success rate of the problems and post-test scores, which was not direct evidence of the explanations' effects. Our study aims at exploring the explanations' effects on learning performance by a direct experiment design, and therefore provides an important contribution to the body of evidence in this field of research.

Mechanism to generate explanations for recommender systems

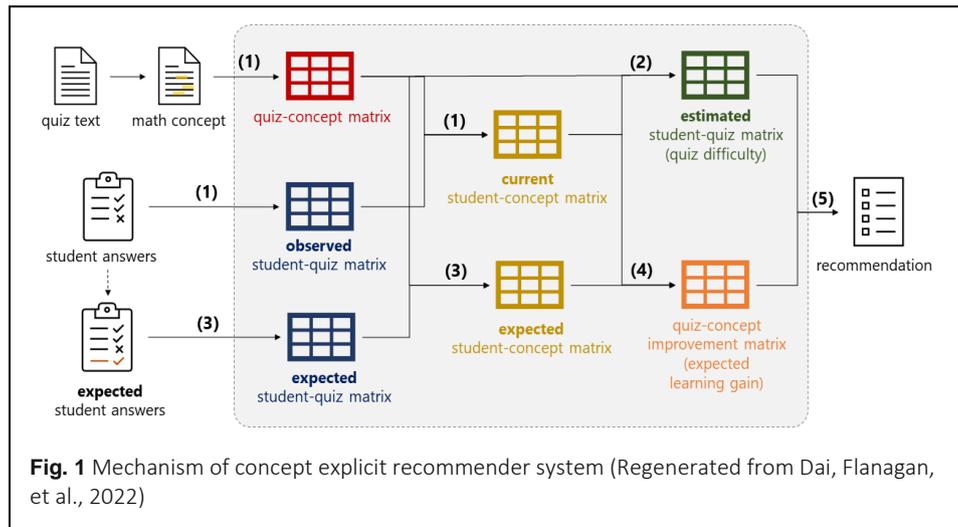
Basically, there are two approaches to generate explanations in recommender systems—model-intrinsic and post-hoc (Zhang & Chen, 2020). In the model-intrinsic approach, the models' mechanism is transparent, and the explanation explains exactly how the model generates a recommendation. To this end, the processes of generating recommendation and generating explanation are mutually dependent. Sometimes, the goal of being explainable constrains the model from being complex and “deep”. In contrast, the post-hoc approach generates the explanation after a recommendation is generated. As a result, the model is allowed to be a “black box” and the explanation does not necessarily explain why an item is recommended. Model-intrinsic approach is desirable when the main purpose is to help users understand why they may need the recommended item. For example, Yu et al. (2021) recommended courses to university students based on the bag-of-words similarities between the candidate course and the courses the students liked. In the explanation, they displayed the common keywords, such as “programming” and “linear”, of the recommended course and the student's favorite course, which reflected how the algorithm worked. Given this explanation, the students can understand the course is recommended because it contains similar topics to the courses s/he is interested in. Other educational recommender systems generate model-intrinsic explanations, such as rule-based (Conati et al., 2021), keyword-based (Yu et al., 2021), concept-based (Dai, Flanagan, et al., 2022; Rahdari et al., 2020), and parameter-based (Takami et al., 2022) methods.

Different types of explanations expose different levels of detailedness of the recommendation algorithm, therefore having different effects on users' perceptions and learning. In this study, we selected the concept-explicit recommender system (Dai, Flanagan, et al., 2022) as the experiment target for the following reasons: 1) the model recommends math quizzes that are supposed to address students' weak concepts, which is considered important in learning math (Birenbaum et al., 1993); 2) The recommender system estimates students' current mastery levels of math concepts and provides visual and textual explanations on this. In learning, knowing what one knows and what one does not know, which is also named as meta-cognitive skill, plays an important role in cognitive learning outcomes (Bahri & Corebima, 2015; Veenman et al., 2014); 3) the explanations reveal how the recommended quizzes can contribute to the students' existing knowledge of math concepts. This is in line with the subsumption theory (Ausubel, 1962), which suggested that a piece of new knowledge can be retained and reused in long-term memory if its relationships with the already-known knowledge can be established. We consider the explanations in the concept-explicit recommender system help students to build the relationships between their current knowledge and the quiz to be solved, and it is anticipated that this will improve their motivation to learn and consolidate what they have learnt.

Concept explicit recommender system

The concept explicit recommender system assumes that solving a math quiz requires the knowledge of related math concepts. For instance, the quiz "*Find the set of all positive divisors of 12*" requires the students know the knowledge of "**set**" and "**positive divisor**". Therefore, a student's ability to solve a quiz depends on his/her mastery level of related math concepts. Again, the student's mastery level of concepts can be improved by attempting new quizzes. Details of the recommender system design were reported in Dai, Flanagan, et al.'s work (2022). In this study, we briefly summarized the mechanism in Figure 1 and the following steps:

- 1) Adopt natural language processing methods to extract math concepts from the quiz texts and compute a quiz-concept matrix whose entries indicate the concept's importance to solve a quiz. Compute a student-quiz matrix where the entries represent students' correctness rates of the quizzes from students' answering histories. Compute the current student-concept matrix which estimates the students' mastery levels of the concepts by multiplying student-quiz and quiz-concept matrices.
- 2) Multiplying student-concept and quiz-concept matrices to estimate the students' probabilities to solve the quizzes successfully. Subtracting this probability from one to acquire the personalized quiz difficulty.



- 3) For each quiz, compute the expected student-concept matrix by assuming the student answers it correctly. Compute student-concept matrix using the expected student-concept matrix and the quiz-concept matrix.
- 4) Compute the expected learning gain, namely, the average improvement of concept mastery level for each quiz by comparing current and expected student-concept matrices.
- 5) Generate recommendations and explanations by utilizing the estimated student-quiz matrix and expected learning gain. The system selects quizzes of appropriate difficulties and ranks the quizzes based on the expected learning gain. Top 5 quizzes and the corresponding explanations are generated and displayed as illustrated in Figure 2, where a panel of current mastery level of math concepts (Recommendation explanation a) is displayed on the top, and a list of recommended quizzes is followed with the corresponding two recommendation explanations. Recommendation explanation b1 converts the estimated difficulty into three levels—high, medium, and low. Recommendation explanation b2 displays the changes of the mastery levels by using the data in 4).

Method

The recommender system was implemented as an application module in the Learning Evidence Analytics Framework (LEAF) (Flanagan & Ogata, 2018), which includes the learning management system Moodle to manage students and course information, the ebook reading application BookRoll to view and answer math quizzes, and the recommender system to view recommendations.

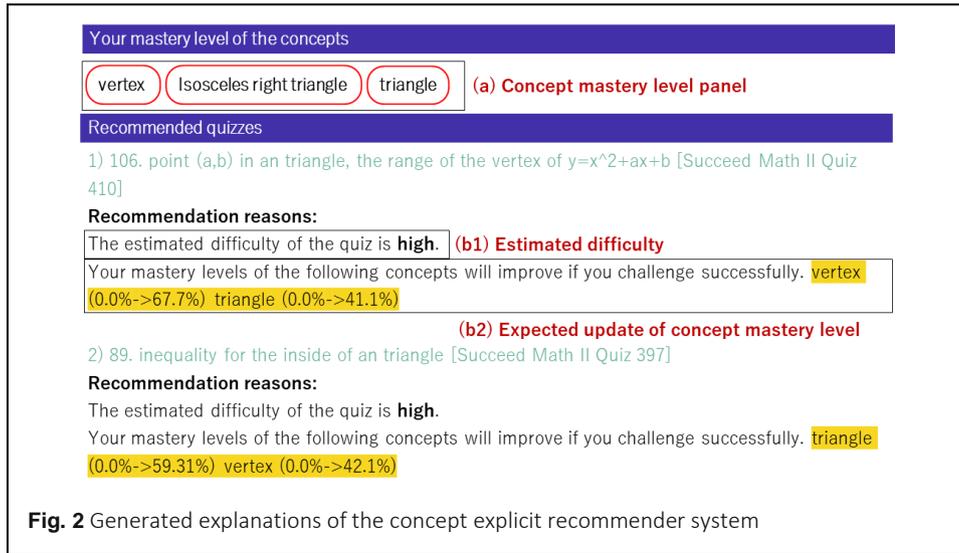


Fig. 2 Generated explanations of the concept explicit recommender system

Experiment design

We conducted a quasi-experiment (Cohen et al., 2011) in a Japanese high school that has adopted the LEAF system as a digital complement to traditional education. 276 first year students (who are generally aged between 15 to 16) whose legal guardian consented to provide their learning data for research use took part in the experiment. We designed the experiment following the procedures outlined in Figure 3.

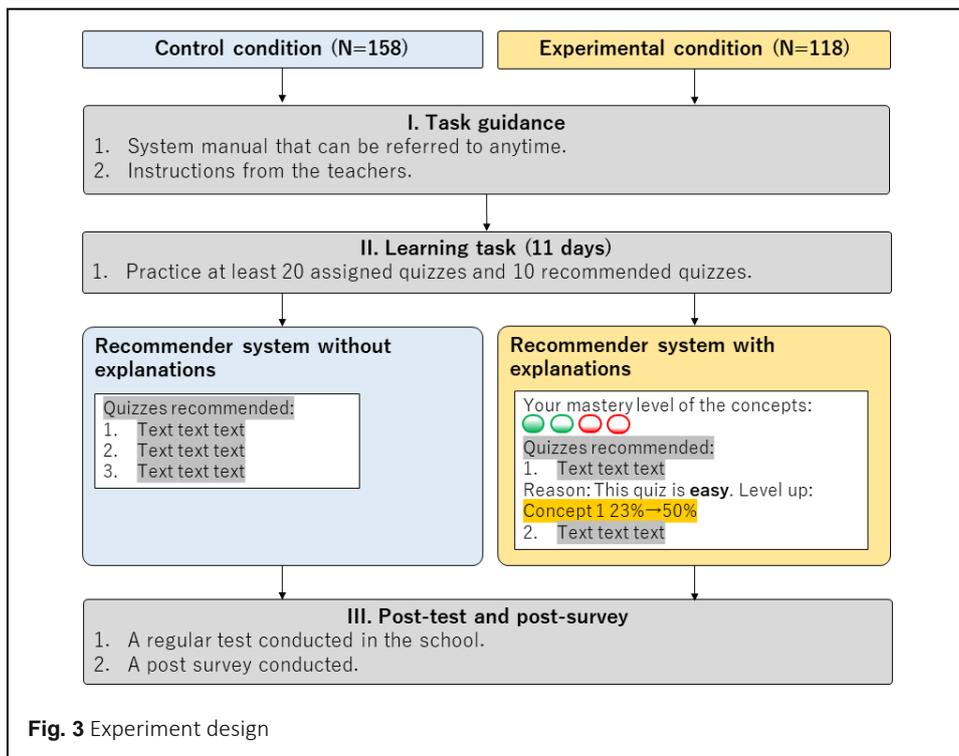


Fig. 3 Experiment design

I. Task guidance.

We prepared a manual to describe how to use the system with detailed screenshots and uploaded it in the course homepage in Moodle. At the same time, we asked the teachers to announce the task and give a brief guidance in their classes. All the students received the same instructions with the minimal information about how to use the recommender system.

II. Learning task.

The students were supposed to solve math quizzes at home in a short holiday preparing for an upcoming regular test on a specific topic “geometry and equation”. The school selected two quiz books for the students to practice and there were 263 quizzes in this topic. The teachers specified 20 quizzes as the assignment and asked the students to additionally complete 10 quizzes recommended by the system. Once the students finish answering a quiz, they were supposed to check their answer with the standard answer and report correct/wrong to the system. Note that the students had both of the hard copies and digital version of the quiz books at hand, and they were free to use either of them to solve the quizzes. However, to get the system’s recommendations, they were strongly encouraged to solve the quizzes and report their answers in the system.

To ensure the ecological validity of the study, we decided to divide the students at a class level at which the usual school activities were conducted. We randomly assigned four classes of students to the control condition where a recommender system without explanations was provided, and three classes of students to the experimental condition where the same recommender system with explanations was provided.

III. Post-test and post-survey.

After the learning task, a school regular test was conducted to assess the students’ understanding of the topic. The test consisted of 7 quizzes pooled from the quiz books and the students were not informed what quizzes would be tested. A post-survey on students’ perceptions of the system was conducted.

Data collection

As shown in Table 2, we collected three types of data—behavioral data, perception data, and assessment data during the experiment. We then explain the details of each data by the method it was collected.

Data collected in the recommender system

Figure 4 illustrates the students’ workflow in the recommender system and the collected log data. 1) To get started, the students have a basic route to access a quiz page in BookRoll: Open the PDF file viewer of the target topic through BookRoll’s directory and then jump between quizzes in different pages using the navigation tools in BookRoll viewer. 2) With the recommender system, the students have two more ways to access a specific quiz: Open

Table 2 Notations and descriptions of collected data

Source	Variable	Description
Log data	viewed_count	The number of times a student viewed a recommended quiz when s/he accessed the recommendation page
	clicked_count	The number of times a student clicked the link of a recommended quiz.
	clickedstats_count	The number of times a student clicked the link of a quiz in the list.
	rec_answered_count	The number of times a student clicked the link of a recommended quiz, solved it, and reported her/his answer.
	answered_count	The number of times a student solved and reported his/her answer to a quiz.
	click_rate	The ratio of clicked recommended quizzes out of viewed recommended quizzes.
	rec_answer_rate	The ratio of answered recommended quizzes out of clicked recommended quizzes.
Survey data (Adapted from Conati et al.'s work (2021))	Q01-> satisfaction	"I was satisfied with the recommender system."
	Q02-> helpfulness	"I think the recommender system is helpful for my math learning."
	Q03-> trust	"I trust the recommender system."
	Q04-> motivation	"I was more motivated to solve the quizzes because of the recommender system."
	Q05-> reason	"I understood why the quizzes were recommended to me."
Assessment data	midterm_score	The score of a midterm regular test in the school.
	posttest_score	The score of a post regular test in the school.
	learning_improvement	The difference between the z-scores of the post-test and the midterm test.

the recommendation page, click the hyperlink of a quiz from the recommendation block on the top or from a quiz list at the bottom. The quiz list lists up all the quizzes in a topic in the default order of the quiz book and shows the students' trial information with marks. This list was added to increase the usability of the recommendation page and the opportunity that students view the recommendations when they prefer to access the quizzes on their own choices. 3) Once the students open a quiz page in BookRoll viewer, they are supposed to solve the quiz using a stylus pen, check the standard answer in the next page, and report their answering situation (correct or wrong) in a quiz tab. If a student reports his/her answer, the system records a quiz answer. We view a quiz answer as a result of the recommendation if there was a previous click on the recommendation hyperlink within one day. Overall, the students can choose to solve the quizzes on their own ways at any phases in the flow.

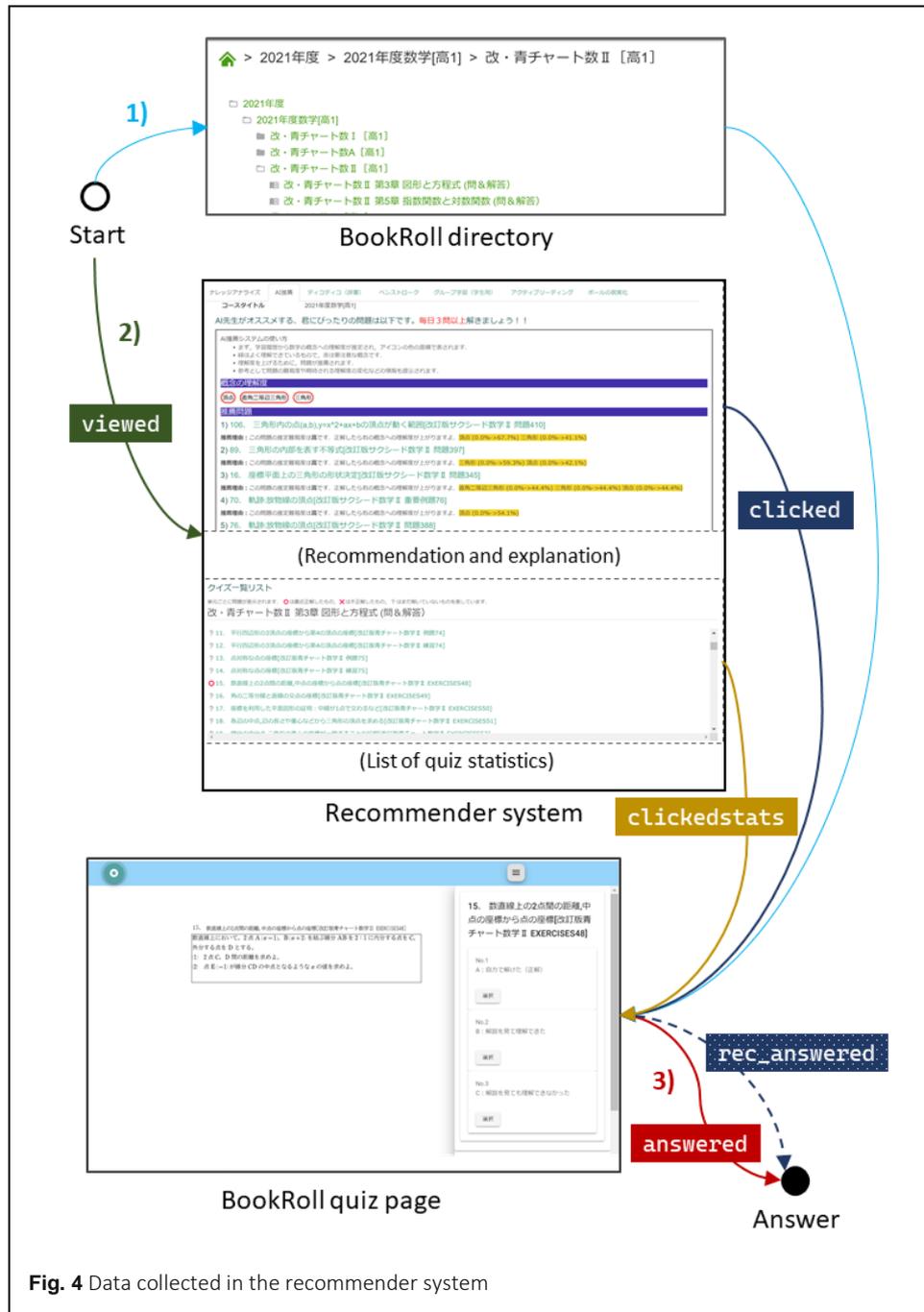


Fig. 4 Data collected in the recommender system

Data collected in the post-survey

To understand students' perceptions on the recommender system, we conducted a short survey which includes five 5-Likert-scale questions after the learning task. The details of each question are shown in Table 2.

Data collected in the regular assessment in the school

A regular assessment on the knowledge of the target topic was scheduled after the experiment period by the school. The test consisted of 7 quizzes selected from the quiz books and was scored from 0 to 100 by the school teachers. Due to the tight learning schedule of the school, we were not able to conduct an extra pre-test to assess students' knowledge of the target topic before using the recommender system. Instead, we utilized a regular midterm test conducted two months before the experiment period as an indicator of students' previous math knowledge. Since the two tests covered different topics, we adopted z-score to normalize the scores and utilize the difference between two z-scores to measure students' improvement in math learning. Note that the z-score difference is not a direct measurement of the learning gain on the target topic but a measurement of students' learning performance improvement among their peers.

Results

Descriptive results

To exclude undesirable usage of the recommender system (e.g., solve the quizzes in hard copies and report answers to the system all at once), we filtered the log data in the sessions where the average time spent on a quiz answer is shorter in one minute (Dai, Takami, et al., 2022). All the data was processed in statistical platform jamovi (The jamovi project, 2022) and Python language (Python Software Foundation, 2022). As Table 3 shows, a part of the students accessed the system and a part of them answered the post-survey. In this study, we consider that the behavioral and assessment data are more objective than perceptual data. As a result, we mainly focused on the students who took both of the tests and had accessed the recommender system in the following analysis (Results 1, 3, and 4). For the analysis of survey data (Result 2), a subset of the students was available and the Cronbach's α was 0.87, which indicates a good reliability.

Table 3 Numbers of students who participated in different activities

Analysis	Group	Control	Experimental
	Total participants	158	118
Results 1, 3, and 4	Students who took both tests and had accessed the recommender system (<i>viewed_count</i> > 0)	81	77
Result 2	Students who took both tests, had accessed the recommender system (<i>viewed_count</i> > 0), and answered the post-survey	30	39

Result 1: Explanations' effects on learning performance

We conducted a student t-test on the learning improvement between two conditions with the hypothesis that the mean learning improvement of the experimental condition is greater than the control condition. To strengthen the persuasiveness, we also conducted a student t-test on the midterm scores between two conditions, which indicates the difference of the prior math ability of the students under two conditions. The results of Levene's test were not significant (learning improvement: $F = 0.005$, $p = 0.943 > 0.05$; midterm score: $F = 2.481$, $p = 0.117 > 0.05$), which indicated the assumption of homogeneity of variance. As Table 4 shows, the t-test revealed a significant difference ($t = -1.671$, $p = 0.048 < 0.05$) in learning improvement, and no significant difference ($t = 0.256$, $p = 0.601 > 0.05$) in the midterm scores of the students under two conditions. These results indicate that the students acquired greater learning improvement when utilizing the recommender system with the explanations, and this was not due to the students had higher or lower levels of prior math ability. Therefore, we positively answer **RQ1**.

Result 2: Explanations' effects on students' perceptions of the recommender system

We conducted student's t-tests on students' perceptions on the system of two conditions. As the Levene's test of *Q02*-> *helpfulness* was significant ($F = 4.231$, $p = 0.044 < 0.05$), we conducted Welch's t test for this variable instead. As Table 5 shows, the mean values of the perceptions of students under the experimental condition were greater than the students under the control condition, and the difference of *Q04*-> *motivation* was statistically significant ($t = -1.668$, $p = 0.05$). As not all the t-test results were significant, we cannot answer **RQ2** positively. However, we did observe higher average values of the perceptions under the experimental condition. We promisingly consider that the explanations in the recommender system positively affect students' perceptions of the system if more data was collected.

Table 4 T-test results of learning improvement and prior math ability of two conditions

	Condition	N	Mean	SD	t	df	p
learning_improvement	control	81	-0.127	0.981	-1.671*	156.00	0.048
	experimental	77	0.133	0.977			
midterm_score (prior math ability)	control	81	55.901	17.897	0.256	156.00	0.601
	experimental	77	55.130	19.931			

Note. $H_a \mu_{experimental} > \mu_{control}$

* $p < .05$

Table 5 T-test results of students' perceptions of two conditions

		Condition	N	Mean	SD	t	df	p
Q01-> satisfaction	Student's t	control	30	3.000	0.983	-1.172	67	0.123
		experimental	39	3.308	1.151			
Q02-> helpfulness	Welch's t	control	30	3.067	0.944	-1.313	67	0.097
		experimental	39	3.410	1.229			
Q03-> trust	Student's t	control	30	2.800	1.064	-0.931	67	0.178
		experimental	39	3.051	1.146			
Q04-> motivation	Student's t	control	30	2.333	0.884	-1.668	67	0.050
		experimental	39	2.769	1.202			
Q05-> reason	Student's t	control	30	3.167	1.234	-1.280	67	0.102
		experimental	39	3.538	1.166			

Note. $H_a \mu_{experimental} > \mu_{control}$

Result 3: Students' behaviors of solving quizzes in the recommender system

We further explored whether there exists a difference between students' recommender system usage under two conditions. As Table 6 shows, interestingly, there were no significant differences between the numbers of viewed recommended quizzes (*viewed_count*), the numbers of clicked recommended quizzes (*clicked_count*), the numbers of clicked quizzes in the quiz list (*clickedstats_count*), the number of answered recommended quizzes (*rec_answered_count*), and the total number of answered quizzes (*answered_count*) under two conditions. Namely, the students under two conditions did not have a difference in terms of the amount of learning. This provides an answer to **RQ3** as the larger learning improvement under the condition of recommender system with explanations is not due to the students solving more quizzes.

The mean values of *answered_count* (control: 27.086, experimental: 28.117) under both conditions met the required numbers of quizzes (total: 30) to solve in the learning task

Table 6 T-test results of students' recommender system usage of two conditions

		Condition	N	Mean	SD	t	df	p
viewed_count	Student's t	control	81	112.778	92.760	1.277	156	0.898
		experimental	77	92.857	103.203			
clicked_count	Student's t	control	81	9.012	7.302	1.553	156	0.939
		experimental	77	7.221	7.190			
clickedstats_count	Student's t	control	81	8.691	9.573	-0.054	156	0.479
		experimental	77	8.779	10.976			
rec_answered_count	Student's t	control	81	7.136	5.850	1.133	156	0.871
		experimental	77	6.052	6.175			
answered_count	Student's t	control	81	27.086	11.751	-0.306	156	0.380
		experimental	77	28.117	27.849			
click_rate	Welch's t	control	81	0.126	0.176	-1.445	137.401	0.075
		experimental	77	0.175	0.245			
rec_answer_rate	Welch's t	control	81	0.628	0.364	-1.385	149.710	0.916
		experimental	77	0.541	0.425			

Note. $H_a \mu_{experimental} > \mu_{control}$

approximately. It can be interpreted as the explanations in the recommender system did not encourage the students to solve more quizzes if the learning task had a specific instruction on how much the students should learn. However, if we further analyze the ratio of clicked recommendations over viewed recommendations (*click_rate*) and the ratio of answered recommendations over clicked recommendations (*rec_answer_rate*), we observed a marginal significant difference of *click_rate* under two conditions. The explanation's effect of motivating students to accept the recommendations was also found in previous studies (Dai, Takami, et al., 2022). In summary, the findings in this study are: 1) the students exposed to the explanations improved more in learning performance, 2) the students exposed to the explanations did not solve more quizzes, and 3) the students exposed to the explanations had a high probability to accept the recommended quizzes. We assume that the explanations may improve the students' determination to solve the quizzes when using the system. In other words, the students do not hesitate on selecting quizzes as the explanations provide clues on how the quiz may contribute to their learning. Without the explanations, the students are not aware of how the quizzes are relevant and merely complete the learning task unwillingly.

Result 4: The relations between students' behaviors of solving quizzes in the recommender system, learning improvement, and students' math abilities

To explore which students' behaviors of solving quizzes in the recommender system may contribute to the learning improvement, we conducted correlation analysis on the students under the experimental condition. First, we examined all the students under the experimental condition, and did not find any significant correlations between the students' behavior indicators (*viewed_count*, *clicked_count*, *clickedstats_count*, *rec_answered_count*, *answered_count*) and *learning_improvement*. Secondly, we decided to investigate if there was a difference based on a student's prior math ability. We divided the students into two groups based on their midterm test z-scores, which was computed among the students who had accessed the system.¹ If the student's z-score was greater than 0, s/he was assigned to the group with high prior math ability. Otherwise, the student was assigned to the group with low prior math ability. As a result, we had a group of students with high math ability (N = 41, Mean = 0.886, SD = 0.574), and a group of students with low math ability (N = 36, Mean = -0.816, SD = 0.438). We then conducted correlation analysis again in each group. Table 7 shows that there is a significant correlation (Pearson correlation coefficient = 0.365, p = 0.014) between the number of viewed recommendations and the learning improvement for the students with low prior math ability. The concept mastery level panel (refer to Explanation.a in Figure 3) and the recommendation reasons under each recommended quiz (refer to Explanation.b1 and Explanation.b2 in Figure 3) is related to the behavior of viewing recommendations. As the students solved more and more

Table 7 Correlations between learning improvement and recommender system usage in the groups of high and low math abilities

	learning_improvement	
	High math ability (N = 41)	Low math ability (N = 36)
viewed_count	-0.044 (p = 0.607)	0.365* (p = 0.014)
clicked_count	0.256 (p = 0.053)	0.073 (p = 0.337)
clickedstats_count	-0.114 (p = 0.762)	0.208 (p = 0.111)
rec_answered_count	0.220 (p = 0.084)	0.044 (p = 0.400)
answered_count	0.127 (p = 0.214)	0.175 (p = 0.154)

Note. H_a is positive correlation

* $p < .05$

quizzes, the better they understand how the explanations changed and how the recommender system function when they viewed the recommendation. However, we did not observe such a correlation in the group of students with high math ability. Instead, there are marginally significant correlations between the number of clicked recommendations and the learning improvement (Pearson correlation coefficient = 0.256, $p = 0.053$), and between the number of answered recommended quizzes and the learning improvement (Pearson correlation coefficient = 0.220, $p = 0.084$). A possible interpretation is that the students may have a higher level of ability to monitor their learning progress. What is more important for these group of students is the action to make a decision on what quizzes to take next time and then execute the decision.

Discussion

Our results show that students achieved more learning improvement by using the recommender system with explanations compared with the students using the recommender system without explanations. In addition, the students using the recommender system with explanations demonstrated more positive perceptions of the system including the satisfaction with the system, the perceived helpfulness of the system, the trust toward the system, the motivation to use the system for math learning, and the understanding of the system's mechanism, though not statistically significant. Besides, Result 3 shows that the students under two conditions did not have a significant difference in terms of the number of solved math quizzes. That is to say, the students solved around 30 quizzes just as required by the teachers. Given this specific learning context where the students had rather high extrinsic motivation to complete the task, the explanations did not encourage the students to attempt and solve more quizzes. However, the explanations may

have a positive effect on the acceptance rate of the recommended quizzes and students' final learning improvement. This finding can be linked to the statements about the functions of open learner models in intelligent tutoring systems. Learner models refer to the learning status that is recorded and maintained by the system (Bull, 2020). By opening the model to the learners, the system can promote learner reflection, facilitate monitoring of learning, preserve learners' rights to access and control over their data, increase the learners' trust towards the content (Bull & Kay, 2013). Our concept-explicit recommender system confronts the students with their learning progress through the concept mastery level panel and allows the students to select the quizzes that fits to their learning needs. As explored in previous research, the meta-cognitive skills—knowing what one knows and controlling one's learning—play an important role in improving cognitive learning outcomes (Bahri & Corebima, 2015; Veenman et al., 2014). Our results imply the possibility that the recommendation explanations affect students' meta-cognitive skills and their perceptions, which further contribute to students' learning improvement. For example, with the concept master level panel, the students become aware of learning goals and their weak concepts. On the contrast, the students may negatively complete the learning tasks without fully convinced that solving recommended quizzes is helpful without the explanations. Consequently, they may access the system more frequently but conduct less actions of attempting the quizzes. To verify this assumption, we need to conduct further studies where meta-cognitive skills are measured directly or indirectly in using the explanations and recommendations in the system.

When we focus on the students under the condition of using the recommender system with explanations, we found that the students with different levels of prior math ability benefit differently from the system. Result 4 shows that for the students with lower math ability, the more they viewed the recommendations, the better they improved learning. Interestingly, there was no significant correlations between the number of clicked recommendations and their learning improvement. A potential interpretation is that viewing the recommendation page exposes the concept mastery level panel to the students, therefore they have a better understanding of their learning progress, which has a positive effect of their overall learning performance. On the other hand, more clicks or answers on the recommended quizzes did not indicate a better learning performance. We consider two possible reasons: 1) The recommendations do not fit to students' mastery level; 2) The students are not able to process the explanations of each individual recommendation and make a good choice of quizzes to solve. For the first reason, we need to investigate the accuracy and validity of the recommendations. For the second reason, we need to further analyze how students make decisions when given the alternatives and explanations. For the students with higher math ability, there was a marginal significant correlation between the number of clicked recommended quizzes and their learning improvement. This implies that

they may benefit from using the recommender system by making more decisions. We consider that the meta-cognitive skills of these students are relatively high, and they may be more critical towards the recommendations and explanations.

Conclusion and future work

In this study, we conducted a comparative experiment ($N = 276$) in high school classrooms, aiming to investigate whether the use of an explainable math recommender system improves students' learning performance. We found that the presence of the explanations had positive effects on students' learning improvement but not the number of solved quizzes during the learning task. The only difference of the recommender systems under two conditions is the presence of the explanations. Although without direct measurement, we interpreted that the overall concept mastery panel, the difficulty and learning gain in the individual explanations help student to track their leaning progress and select the quizzes at their own choices, which affects students' meta-cognitive skills and their perceptions. These meta-cognitive skills and perceptual elements further contribute to the learning improvement (Bahri & Corebima, 2015; Veenman et al., 2014). When separating the students based on their prior math abilities, we found a significant correlation between the number of viewed recommendations and the final learning improvement for the students with lower math abilities. This indicates that the current recommendation explanations fit to the students with lower math abilities. For students with higher math abilities, more sophisticated and interactive recommender system is more desirable. The main contributions of this study are three-fold:

- 1) We conducted a continuous real-life experiment in high school classes to investigate the recommendation explanations' effects on learning performance, which provides important empirical evidence to this research field.
- 2) We found that the recommendation explanations positively affected the learning improvement without leading to more attempts on the quizzes. As the recommendation explanations revealed students' learning progress and provided opportunities of allowing the students to make their own choices of quizzes to solve, the explanation may positively affect students' meta-cognitive skills, which then efficiently affect learning performance.
- 3) We found that students with different levels of math abilities had different behaviors in the recommender system, which implies a necessity to personalize the recommendations and explanations based on students' characteristics.

There are still some limitations in this study. In this study, we mainly focused on the effects of the recommendation explanations. Since the explanations of the concept-explicit recommender system are model-intrinsic, the goodness of the explanations is intertwined with the goodness of the recommendations. In the future work, more investigation should

be conducted to reveal how the recommender system can help learning as a whole. Regarding students' perceptions and reactions on the explanations, we need to record more log data or resort to other methods such as interviews to understand which part of the explanations causes a specific action, and how it is related to meta-cognitive skills. As described in the discussion, one of other future directions is to improve the recommender system with personalized or interactive explanations for advanced students.

Abbreviations

LEAF: Learning Evidence Analytics Framework

Endnotes

¹ A larger number of groups was not feasible as it resulted in insufficient samples in each group.

Authors' contributions

YD implemented the system, designed the experiment, performed data analysis and drafted the initial manuscript. KT co-conducted the experiment and edited the manuscript. BF provided the support to implement the system and edited the manuscript. HO was responsible for funding acquisition and supervision. All authors read and approved the final manuscript.

Authors' information

Yiling Dai is a Program-Specific Researcher at the Academic Center for Computing and Media Studies, Kyoto University. She received a bachelor's degree from Zhejiang University, a master's degree from the Graduate School of Business, Rikkyo University, and a PhD degree from the Graduate School of Informatics, Kyoto University. Her research interests include: Information Retrieval, Knowledge Discovery, Educational Data Mining and Learning Analytics.

Kyosuke Takami is a senior researcher at Education Data Science Center, National Institute for Educational Policy Research (NIER) in Japan. He has prior expertise in the application of neuroscience in education and received a PhD degree from Osaka University. He has been a high school teacher for more than 9 years. His research interests include: Learning Analytics, Educational Data Mining and Educational Technologies.

Brendan Flanagan is an Associate Professor at the Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, and the Graduate School of Informatics at Kyoto University. He received a bachelor's degree from RMIT University and master's and Ph.D. degrees from the Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests include: Learning Analytics, Educational Data Science, Educational Data Mining, NLP/Text Mining, Machine Learning, Computer Assisted Language Learning, and the Application of Blockchain in Education.

Hiroaki Ogata is a full Professor at the Academic Center for Computing and Media Studies, Kyoto University. His research interests include: Learning Analytics, Evidence-Based Education, Educational Data Mining, Educational Data Science, Computer Supported Ubiquitous and Mobile Learning, and CSCL.

Funding

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) JP20H01722 and JP23H01001, (Exploratory) JP21K19824, (Early Career) JP23K17012, (A) JP23H00505, and NEDO JPNP20006.

Availability of data and materials

Not applicable.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan.

² Education Data Science Center, National Institute for Educational Policy Research, Tokyo, Japan.

³ Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, Kyoto University, Kyoto, Japan.

Received: 14 November 2022 Accepted: 4 September 2023

Published online: 1 January 2024 (Online First: 12 September 2023)

References

- Ausubel, D. P. (1962). A subsumption theory of meaningful verbal learning and retention. *The Journal of General Psychology*, 66(2), 213–224. <https://doi.org/10.1080/00221309.1962.9711837>
- Bahri, A., & Corebima, A. D. (2015). The contribution of learning motivation and metacognitive skill on cognitive learning outcome of students within different learning strategies. *Journal of Baltic Science Education*, 14(4), 487–500. <https://doi.org/10.33225/jbse/15.14.487>
- Barria-Pineda, J., Akhuseyinoglu, K., Želem-Čelap, S., Brusilovsky, P., Milicevic, A. K., & Ivanovic, M. (2021). Explainable recommendations in a personalized programming practice system. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin & V. Dimitrova (Eds.), *Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science, vol 12748* (pp. 64–76). Springer, Cham. https://doi.org/10.1007/978-3-030-78292-4_6
- Bilgic, M., & Mooney, R. J. (2005). Explaining recommendations: Satisfaction vs. promotion. In R. St. Amant, J. Riedl & A. Jameson (Eds.), *Proceedings of Beyond Personalization 2005: A Workshop on the Next Stage of Recommender Systems Research at the 2005 International Conference on Intelligent User Interfaces* (pp. 13–18). Association for Computing Machinery.
- 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.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. <https://doi.org/10.1016/j.knsys.2013.03.012>
- Bull, S. (2020). There are open learner models about! *IEEE Transactions on Learning Technologies*, 13(2), 425–448. <https://doi.org/10.1109/TLT.2020.2978473>
- Bull, S., & Kay, J. (2013). Open learner models as drivers for metacognitive processes. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 349–365). Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_23
- Cohen, L., Manion, L., & Morrison, K. (2011). Experiments, quasi-experiments, single-case research and internet-based experiments. In L. Cohen, L. Manion & K. Morrison (Eds.), *Research methods in education (7th ed.)* (pp. 312–334). Routledge.
- Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298, 103503. <https://doi.org/10.1016/j.artint.2021.103503>
- Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5), 455–496. <https://doi.org/10.1007/s11257-008-9051-3>
- da Silva, F. L., Slodkowski, B. K., da Silva, K. K. A., & Cazella, S. C. (2023). A systematic literature review on educational recommender systems for teaching and learning: Research trends, limitations and opportunities. *Education and Information Technologies*, 28(3), 3289–3328. <https://doi.org/10.1007/s10639-022-11341-9>
- Dai, Y., Flanagan, B., Takami, K., & Ogata, H. (2022). Design of a user-interpretable math quiz recommender system for Japanese high school students. In B. Flanagan, R. Majumdar, H. Li, A. Shimada, F. Okubo & H. Ogata (Eds.), *Proceedings of the 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior - DC in LAK22 Co-Located with 12th International Learning Analytics and Knowledge Conference (LAK22)* (pp. 30–38). CEUR Workshop Proceedings.
- Dai, Y., Takami, K., Flanagan, B., & Ogata, H. (2022). Investigation on practical effects of the explanation in a K-12 math recommender system. In S. Iyer et al. (Eds.), *Proceedings of the 30th International Conference on Computers in Education* (pp. 7–12). Asia-Pacific Society for Computers in Education.
- Erdt, M., Fernández, A., & Rensing, C. (2015). Evaluating recommender systems for technology enhanced learning: A quantitative survey. *IEEE Transactions on Learning Technologies*, 8(4), 326–344. <https://doi.org/10.1109/TLT.2015.2438867>
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning*, 10(4), 469–484. <https://doi.org/10.34105/j.kmel.2018.10.029>
- Huang, X., Fang, Q., Qian, S., Sang, J., Li, Y., & Xu, C. (2019). Explainable interaction-driven user modeling over knowledge graph for sequential recommendation. In L. Amsaleg, B. Huet, M. Larson, G. Gravier, H. Hung, C.-W. Ngo & W. T. Ooi (Eds.), *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 548–556). Association for Computing Machinery. <https://doi.org/10.1145/3343031.3350893>
- Karga, S., & Satratzemi, M. (2019). Using explanations for recommender systems in learning design settings to enhance teachers' acceptance and perceived experience. *Education and Information Technologies*, 24(5), 2953–2974. <https://doi.org/10.1007/s10639-019-09909-z>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In J. Kaye, A. Druin, C. Lampe, D. Morris & J. P. Hourcade (Eds.), *Proceedings of the 2016 CHI Conference on Human Factors in*

- Computing Systems* (pp. 2390–2395). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858402>
- Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, 74, 12–32. <https://doi.org/10.1016/j.dss.2015.03.008>
- Murgia, E., Landoni, M., Huibers, T., Fails, J. A., & Pera, M. S. (2019). The seven layers of complexity of recommender systems for children in educational contexts. In M. Koolen, T. Bogers, B. Mobasher & A. Tuzhilin (Eds.), *Proceedings of the Workshop on Recommendation in Complex Scenarios Co-Located with 13th ACM Conference on Recommender Systems (RecSys 2019)* (pp. 5–9). CEUR Workshop Proceedings.
- Python Software Foundation. (2022). *Python Language Reference* (3.8.8) [Computer software]. <http://www.python.org>
- Rahdari, B., Brusilovsky, P., Thaker, K., & Barria-Pineda, J. (2020). Using knowledge graph for explainable recommendation of external content in electronic textbooks. In S. Sosnovsky, P. Brusilovsky, R. Baraniuk & A. Lan (Eds.), *Proceedings of the Second International Workshop on Intelligent Textbooks 2020 Co-Located with 21st International Conference on Artificial Intelligence in Education (AIED 2020)* (pp. 50–61). CEUR Workshop Proceedings.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Takami, K., Dai, Y., Flanagan, B., & Ogata, H. (2022). Educational explainable recommender usage and its effectiveness in high school summer vacation assignment. In A. F. Wise, R. Martinez-Maldonado & I. Hilliger (Eds.), *Proceedings of the 12th International Learning Analytics and Knowledge Conference* (pp. 458–464). Association for Computing Machinery. <https://doi.org/10.1145/3506860.3506882>
- The jamovi project. (2022). *Jamovi* (2.3) [Computer software]. <https://www.jamovi.org/>
- Veenman, M. V. J., Bavelaar, L., Wolf, L. D., & Haaren, M. G. P. V. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences*, 29, 123–130. <https://doi.org/10.1016/j.lindif.2013.01.003>
- Wang, G., Zhao, J., Van Kleek, M., & Shadbolt, N. (2022). Informing age-appropriate AI: Examining principles and practices of AI for children. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson & K. Yatani (Eds.), *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–29). Association for Computing Machinery. <https://doi.org/10.1145/3491102.3502057>
- Wang, N., Wang, H., Jia, Y., & Yin, Y. (2018). Explainable recommendation via multi-task learning in opinionated text data. In K. Collins-Thompson, Q. Mei, B. Davison, Y. Liu & E. Yilmaz (Eds.), *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 165–174). Association for Computing Machinery. <https://doi.org/10.1145/3209978.3210010>
- Wang, X., Chen, Y., Yang, J., Wu, L., Wu, Z., & Xie, X. (2018). A reinforcement learning framework for explainable recommendation. In *Proceedings of 2018 IEEE International Conference on Data Mining (ICDM)* (pp. 587–596). The Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICDM.2018.00074>
- Xian, Y., Fu, Z., Muthukrishnan, S., de Melo, G., & Zhang, Y. (2019). Reinforcement knowledge graph reasoning for explainable recommendation. In B. Piwowarski, M. Chevalier, E. Gaussier, Y. Maarek, J.-Y. Nie & F. Scholer (Eds.), *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 285–294). Association for Computing Machinery. <https://doi.org/10.1145/3331184.3331203>
- 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. <https://doi.org/10.1007/s40747-021-00315-y>
- Yu, R., Pardos, Z., Chau, H., & Brusilovsky, P. (2021). Orienting students to course recommendations using three types of explanation. In J. Masthoff, E. Herder, N. Tintarev & M. Tkalčič (Eds.), *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 238–245). Association for Computing Machinery. <https://doi.org/10.1145/3450614.3464483>
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1), 1–101. <https://doi.org/10.1561/1500000066>

Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research and Practice in Technology Enhanced Learning (RPTEL)
is an open-access journal and free of publication fee.