

RESEARCH

Open Access



Personality-based tailored explainable recommendation for trustworthy smart learning system in the age of artificial intelligence

Kyosuke Takami^{1,2*} , Brendan Flanagan³, Yiling Dai¹ and Hiroaki Ogata¹

*Correspondence:
takami@nier.go.jp

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

² Present Address: 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

Abstract

In the age of artificial intelligence (AI), trust in AI systems is becoming more important. Explainable recommenders, which explain why an item is recommended, have recently been proposed in the field of learning technology to improve transparency, persuasiveness, and trustworthiness. However, the methods for generating explanations are limited and do not consider the learner's cognitive perceptions or personality. This study draws inspiration from tailored intervention research in public health and investigates the effectiveness of personality-based tailored explanations by implementing them for the recommended quizzes in an explainable recommender system. High school students ($n = 217$) were clustered into three distinct profiles labeled Diligent ($n = 77$), Fearful ($n = 72$), and Agreeable ($n = 68$), based on the Big Five personality traits. The students were divided into a tailored intervention group ($n = 106$) and a control group ($n = 111$). In the tailored intervention group, personalized explanations for recommended quizzes were provided based on student profiles, with explanations based on quiz characteristics. In the control group, only non-personalized explanations based on quiz characteristics were provided. An 18-day A/B experiment showed that the tailored intervention group had significantly higher recommendation usage than the control group. These results suggest that personality-based tailored explanations with a recommender approach are effective for e-learning engagement and imply improved trustworthiness of AI learning systems.

Keywords: Tailored intervention, Big five personality traits, Explainable recommendation, K-12 mathematics

Introduction

In the age of artificial intelligence, trust is crucial in development and acceptance of AI (Siau & Wang, 2018). Trust in technology is determined by human characteristics such as personality (Hengstler et al., 2016). In school settings, teachers modify their instructional methods according to the individual personalities of their students. If such personality-based interventions are implemented in AI learning systems, trustworthiness and acceptance of AI may be increased.

Explainable recommenders, which explain why an item is recommended, have recently been proposed in the field of learning technology (Barria-Pineda et al., 2021; Rahdari

et al., 2020; Takami et al., 2022a; Tsiakas et al., 2020). They can improve transparency, persuasiveness, and trustworthiness (Zhang & Chen, 2020). Examples include explanations of learning history, difficulty, or relevance of knowledge in recommended quizzes. However, these explanations do not consider how learners perceive them, or what kind of explanation is best for a learner's characteristics or personality.

In public health research, tailored interventions designed to reach a specific person based on their unique characteristics have been shown to be effective for behavioral change (Sohl & Moyer, 2007). Tailored interventions use individually focused messages delivered by a person, letter, or computer (Kreuter et al., 1999). Previous research suggests that tailored messages may affect people differently (Sohl & Moyer, 2007). Tailored interventions have been studied extensively in public health but have not been fully considered in technology-enhanced education.

This study focuses on a math-quiz recommender system in which quiz-characteristic-based explanations are provided to motivate students to accept recommended quizzes. We hypothesized that additional profile-specific explanations would influence student perceptions of the recommended quizzes, increasing their engagement with the recommender. Although profile-specific persuasive explanations are generated independently of how the recommendation is made, they reveal personality-related information about the recommended item. We conducted an A/B experiment to examine the effectiveness of personality-based tailored interventions in educational recommenders, comparing personality-based tailored explanations in the intervention group and quiz-characteristic explanations in the control group.

Related work

Tailored intervention

Tailored interventions are designed to reach individuals based on their unique characteristics and have shown promise in public health research, such as promoting mammography (Rimer et al., 1999). Tailored interventions include assessment-based, individually focused messages (Kreuter et al., 1999). The assessment involves a closed-ended measure of individual differences. This enables a message tailored to an individual's answers to be pre-established. This scripted message can be delivered by a person (not necessarily a health professional), letter, or computer. Interventions are tailored to a variety of characteristics including age, ethnicity, risk, and barriers to care, or according to theoretical models such as the Health Belief Model, the Transtheoretical Model, and concepts related to motivational interviewing. A meta-analysis review reported that tailored interventions, particularly those that use the Health Belief Model, are effective in promoting mammography screening (Sohl & Moyer, 2007). A tailored intervention approach has recently been initiated in the field of learning analytics (Matz et al., 2021; Tempelaar et al., 2021). Matz et al. proposed tailored support using student profiles of learning style but not personality traits; research in this area is limited.

Persuasion

Persuasive communication intends to change, reinforce, or shape another person's response(s) (Cialdini, 2001; Fogg, 2002). One of the most influential models of persuasive strategies was presented by Cialdini (2001) and included six principles: authority,

consensus, commitment, scarcity, liking, and reciprocity. Authority is considered as a form of social influence; it is suggested that people are inclined to follow suggestions and recommendations from a person with authority (Blass, 1991; Milgram & van Gasteren, 1974), Commitment refers to the notion that people strive to maintain consistent beliefs and act in accordance with those beliefs (Cialdini, 2001). Liking refers to the tendency of people to say 'yes' to people they like (Cialdini, 2001).

Personality traits and persuasion

Personality inventories are psychological questionnaires that reveal the personality traits of participants to better understand their behavior in different settings. The Big Five Inventory (John et al., 1991) describes an individual's personality across five dimensions: openness to experience (O), extraversion (E), agreeableness (A), conscientiousness (C), and neuroticism (N). Previous Big Five personality studies in education have shown relationships between the Big Five dimensions and learning styles (Busato et al., 1998), academic success (O'Connor & Paunonen, 2007), and academic dishonesty (Giluk & Postlethwaite, 2015). A previous questionnaire-based personality and persuasion study reported that people who were agreeable tended to be persuaded by people they like (Alkış & Taşkaya Temizel, 2015), whereas people who were conscientious tended to be persuaded by people with authority (Alkış & Taşkaya Temizel, 2015). Authority is considered a form of social influence and posits that people are inclined to follow recommendations from the person in authority (Blass, 1991; Milgram & van Gasteren, 1974). Fearful individuals are more susceptible to commitment strategies (Wall et al., 2019). Commitment refers to the notion that individuals strive to maintain consistent beliefs (Cialdini, 2001). These traits can be clustered into three types (Asendorpf, 2002; Robins et al., 1996). Based on these findings, we hypothesized that an intervention that added profile-specific explanations would increase engagement.

Explainable recommendation

Explainable recommendations, which explain why an item is recommended, can improve transparency, persuasiveness, and trustworthiness (Zhang & Chen, 2020). Although explainable recommendation research has been conducted mainly in e-commerce, including Amazon and Netflix (Nunes & Jannach, 2017), there has also been a growing interest in explainable recommendation research in the field of education (Barria-Pineda et al., 2021; Dai et al., 2023; Rahdari et al., 2020; Takami et al., 2022a; Tsiakas et al., 2020). Examples include cognitive training for elementary children (Tsiakas et al., 2020), mathematics in high school (Dai et al., 2023; Takami et al., 2022a), personalized programming practice systems in higher education (Barria-Pineda et al., 2021), and Wikipedia article recommendations for online electronic textbook users (Rahdari et al., 2020). These systems are not only for making recommendations; they also generate explanations as to why the recommendations are being made. Explanations of recommended items can be generated from different data sources and provided in different display styles (Tintarev & Masthoff, 2015) including a relevant user or item, a sentence, an image, or a set of reasoning rules. There are two ways to generate explanations in recommender systems: model-intrinsic and post hoc approaches (Zhang & Chen, 2020). In the model-intrinsic approach, the model mechanism is transparent, and the explanation

explains exactly how the model generates recommendations. The post hoc approach generates the explanation after a recommendation is generated (providing simple statistical information such as ‘70% of your friends bought this item’). Post hoc explanations are not invalid, they are simply decoupled from the model.

In either approach, the generated explanation is related to how the algorithm selects the item and why it considers the item to be important to the user, such as increasing knowledge in the case of e-learning. Thus, many explanations in educational recommender systems focus on the characteristics of the learning materials. For instance, explanations of how recommended learning tasks can improve student understanding of prerequisites or key concepts (Barria-Pineda et al., 2021; Dai et al., 2022; Rahdari et al., 2020), how the recommended courses are related to student background and interests in terms of the topics they cover (Yu et al., 2021), and how the learning performance score is predicted (Conati et al., 2021) have been proposed in previous research. However, these studies did not consider how students perceived the recommendations and explanations.

This study used a post hoc approach to generate tailored persuasive explanations from simple statistical information (such as how many top achievers solved recommended quizzes, how many solved today’s task, and how many tasks your classmates solved), and examined the effectiveness of tailored intervention in the educational field. The following research question was addressed:

RQ: Is personality-based tailored intervention effective in an educational explainable recommender system?

Method

Learning data collection

In this study, a personalized explainable recommender was developed on a learning system designed to support the distribution of learning materials and the collection and automated analysis of learning behavior logs using an open, standards-based approach (Flanagan & Ogata, 2018). The overall architecture of the system is shown in Fig. 1. The

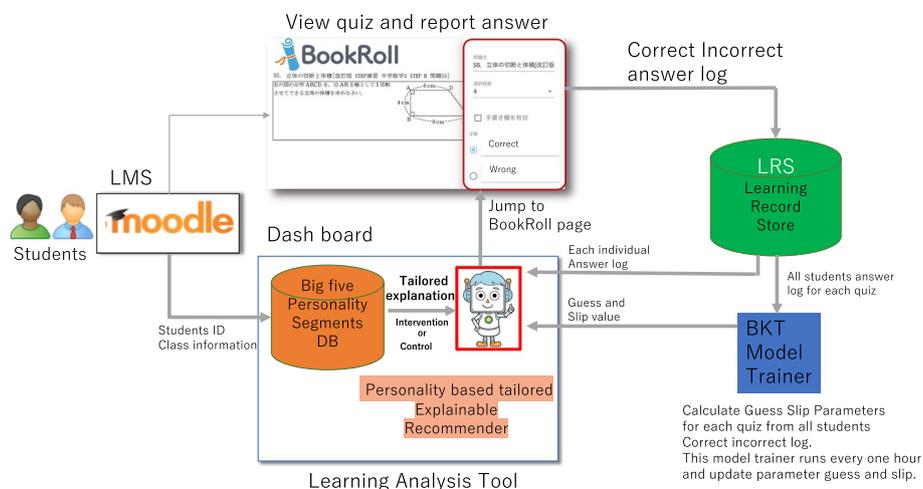


Fig. 1 Overall architecture of personality-based tailored explainable recommender

main components of the framework are the Moodle LMS, which acts as a hub for accessing different courses, the BookRoll reading system for learning material and quiz exercise distribution, an LRS for collecting learning behavior logs from all components, and the learning analytics dashboard to provide feedback to students, teachers, and school administrators. This framework enabled us to collect and analyze learning behaviors in real time and provide feedback to stakeholders. The quiz books used in mathematics classes were uploaded to the reading system and multiple-choice quiz questions were created to enable collection of answers in the learning log data. We collected log data, including student-accessed data, quiz-clicked data, and answered data on the quizzes, right or wrong.

Explainable recommender

From the collected right or wrong quiz data, recommendation of learning paths was calculated using Bayesian knowledge tracing (BKT) (Corbett & Anderson, 1994) to model the degree of mastery of each skill (quiz) in the recommender based on analysis of answers in the learning log data using the Python Library of the BKT model (Badri-nath et al., 2021). Quizzes were recommended based on the probability that the student would correctly answer a question, as determined by the BKT model, with quizzes with an extremely high or low probability of correct answers having less weight in the recommendation (Fig. 1, right).

As a basic common control explanation, we used an explanation generator using the BKT parameter guess (giving a correct answer despite not knowing the skill) and slip (knowing a skill but giving a wrong answer) (Takami et al., 2021). The explanation generator categorized recommended quizzes into different feature types according to the values of the model parameters and output explanation texts (i.e., High Guess value, meaning new skills: “Let’s carefully go over some basic skills with this problem!”; Low Guess value, meaning previous skills required: “Let’s try this quiz! This is a quiz that you can solve using your learned skills.”; High Slip value, meaning careless mistakes: “This quiz is so easy to miss!”, etc.) based on the categorized feature types. In this study, the explanation of quiz characteristics generated from BKT parameters was used as a common baseline. The intervention group generated additional tailor-made explanations along with this explanation.

Participants

Ten high school mathematics classes participated in this study. We obtained consent from all participants for their cooperation in this study and for use of their learning logs in our research. At the beginning of the semester, students were divided into science and humanities courses according to their career aspirations, and into proficiency classes (*advanced*, *standard*, and *basic*) based on their grades, with classes of the same proficiency level categorized into approximately the same academic level, as shown in Table 1. For example, the advanced classes A and B were adjusted for the same academic ability group. During the experimental period, all classes studied the same course content (“vectors of planes”), used the same teaching materials, and progressed in learning in the same manner, with slight differences in the way the lessons were explained according to the academic ability of the students. There were 114 quizzes on the studied content; of

Table 1 Math classes involved in intervention experiment

Classes	Enrollment	Participant	Diligent	Fearful	Agreeable	Intervention/Control
Science course <i>advanced</i> Class A	40	37	23	7	7	Control
Science course <i>advanced</i> Class B	40	33	15	12	6	Intervention
Science course <i>standard</i> Class C	27	22	9	9	4	Control
Science course <i>standard</i> Class D	26	18	8	4	6	Intervention
Science course <i>basic</i> Class E	13	9	2	4	3	Control
Science course <i>basic</i> Class F	13	11	2	5	4	Intervention
Humanity course <i>advanced</i> Class G	39	31	8	10	13	Control
Humanity course <i>standard</i> Class H	39	29	5	11	13	Intervention
Humanity course <i>standard</i> Class I	20	12	3	3	6	Control
Humanity course <i>standard</i> Class J	19	15	2	7	6	Intervention
Total	276	217	77	72	68	

The class names in bold indicate the intervention group

these, five quizzes were recommended based on the level of understanding calculated using the BKT model. The teacher directed the students to use the recommender system, solve the quizzes, and check their answers on the system. Before the experiments, we used the Big Five personality questionnaire (John et al., 1991) to measure personality traits (O, A, C, E, and N) of second-year high school students on a 12-point scale. Personality data with no missing values were obtained from the 217 students. In addition, we asked the students whether they were good at math, based on a five-point Likert scale. Math anxiety (Luttenberger et al., 2018) is a major problem in mathematics; we thought that whether students were good at mathematics may have a great influence on persuasiveness. Previous studies have shown that self-assessments of mathematics perception are related to past academic performance (Hackett & Betz, 1989; Lopez & Lent, 1992). The personality data (Big Five and math self-assessment) were stored individually for each student in the personality segment database (Fig. 1, bottom).

Clustering student personality

We clustered the Big five personality traits and good-at-math scales using k-means clustering. Figure 2 shows an elbow plot that indicates the transition of cluster information over cluster numbers from 1 to 10. From the figure, we can observe that the optimal cluster number is three based on the resemblance to an elbow at this point. This result is consistent with previous results that classified personality traits into three types (Asendorpf, 2002; Robins et al., 1996; Wall et al., 2019). Figure 3 shows the mean scores of the profiling personality variables (12-point scale) and the math self-assessment (5-point scale). Profile 1 ($n=77$, 35.5% of the sample), labelled as 'Diligent', comprised individuals who reported greater Openness and Conscientiousness, and were good at math. Profile 2 ($n=72$, 33.2% of the sample), labelled as 'Fearful', reported higher levels of Neuroticism.

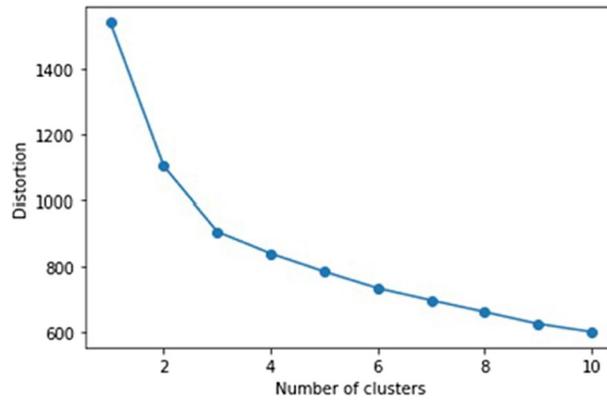


Fig. 2 Elbow plot

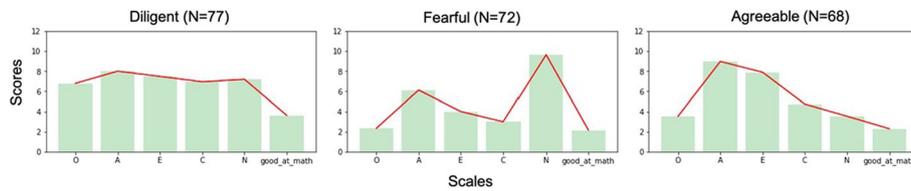


Fig. 3 Mean of personality variables for each profile

Table 2 Personality trait statistics for each profile

Personality variables	Diligent (N = 77)	Fearful (N = 73)	Agreeable (N = 69)	Univariate <i>F</i> (2, 214)
Openness (O)	6.792 (2.577)	2.333 (2.182)	3.529 (2.560)	66.548***
Agreeableness (A)	8.000 (2.575)	6.125 (2.696)	8.971 (2.311)	22.928***
Extraversion (E)	7.481 (3.556)	4.000 (3.525)	7.897 (3.447)	26.606***
Conscientiousness (C)	6.948 (2.590)	2.986 (1.954)	4.735 (2.410)	53.793***
Neuroticism (N)	7.195 (3.449)	9.639 (2.309)	3.529 (2.788)	78.246***
Good at math	3.623 (0.918)	2.153 (0.944)	2.294 (0.947)	56.113***

Personality variables: mean, standard error in parentheses, and mean difference across three personality clusters, *** $p < 0.001$

Respondents in Profile 3 ($n = 68$, 31.3% of the sample), labelled as ‘Agreeable’, reported higher levels of Agreeableness and Extraversion. Information regarding these three segments was assigned and accumulated for each individual in the personality segment database (Fig. 1, bottom).

A one-way MANOVA revealed significant differences in the profiling personality variables ($F(2, 214) = 55.675$, $p < 0.001$; Wilk’s lambda = 0.148). Post hoc comparisons, summarized in Table 2, revealed that individuals in the Diligent group reported significantly

higher scores for Openness, Conscientiousness, and math proficiency than those in the Fearful and Agreeable groups. Individuals in the Fearful group reported significantly higher Neuroticism than those in other groups. Agreeable individuals tended to have higher scores for agreeableness than for other traits.

Figure 4 shows the flow of the participants. A total of 217 students labelled in one of the three profiles were assigned to either a tailored intervention group with a personality-based explanation (n = 106) or a control group (n = 111) without a personality-based explanation. Allocation of intervention and control groups was done in such a way that classes with similar levels of academic ability were effectively separated, as shown in Table 1. For instance, in the advanced-level classes, classes A and B were assigned to the control and intervention groups, respectively. Additionally, consideration was given to ensuring a broad distribution of personality types. Class information was sent as session information from Moodle and tailor-made explanations were generated for each segment using the profile segment database (Fig. 1, left). In these conditions, an explainable recommender was available for 18 days, from May 8–25, 2022, to implement the experimental contrast as an A/B test (tailored intervention group versus control group). During the experimental period, the log data of the accessed system, clicks on recommendations, and clicks on the quiz-stats list were collected. There were no significant differences in each personality scale between the intervention and control groups, except in Cluster 1 for the good-at-math scale (mean 1.872 and 2.485, SD 0.864 and 0.983, $t = -2.882$, $df = 70$, $p = 0.005$, comparing the intervention group to the control group).

Personality-based tailored explanation

Three types of tailored persuasive explanation suited to each profile were developed through previous personality and persuasion studies (conscientious individuals tend to be persuaded by people with authority (Alkış & Taşkaya Temizel, 2015), fearful individuals are more susceptible to the commitment strategy (Wall et al., 2019), and agreeable individuals tend to be persuaded by people they like (Alkış & Taşkaya Temizel, 2015). Table 3 shows examples of each tailored persuasive statement. For the Diligent profile, with high openness and conscientiousness, authority-related explanations of how many

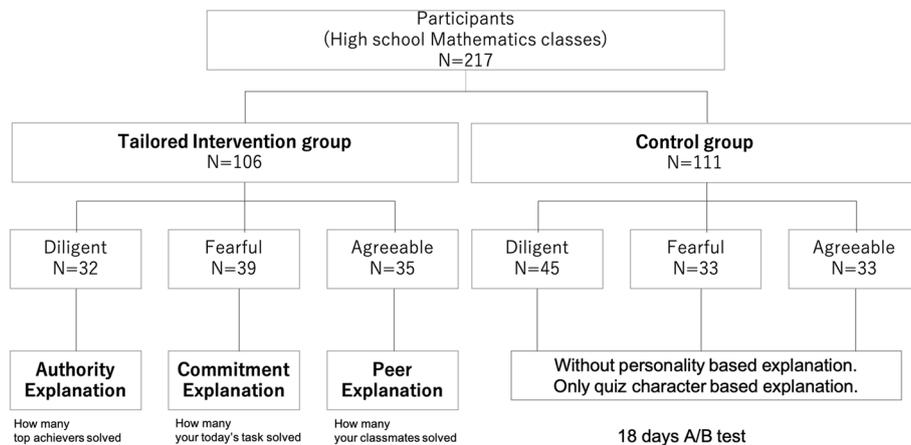


Fig. 4 Flow diagram of experiment design

Table 3 Examples of each tailored persuasive statement

Profile	Persuasive type	Type of explanation	Texts of explanations
Diligent	Authority	How many top achievers	“Three top achievers have solved this quiz.” “None of the top achievers have solved this quiz yet.”
Fearful	Commitment	How many of daily tasks	“Today you used one of the recommended quizzes. Two more quizzes and you will have achieved your daily goal.” “You did it! Today you used four of the recommended quizzes. You’ll solve more of these recommended quizzes to further solidify your understanding!”
Agreeable	Peer	How many of your classmates	“This quiz was solved by two students in your class. Let’s solve yours!” “No one in the class has solved this quiz yet.”
Control			Only quiz-character-related explanations



Fig. 5 Screenshot of recommendation UI

top achievers solved the quiz were provided. For the Fearful profile with high neuroticism, commitment-related explanations were displayed indicating how many of today’s tasks were completed. Peer-related explanations were provided for the Agreeable profile. In the control group, only quiz-character-related explanations were provided, also displayed in the profile-based explanation group.

User interface

We implemented tailored persuasive explanation in the explainable recommendation system. Figure 5 shows a screenshot of the user interface of the explainable recommendation system. A quiz feature-based explanation of the recommendations was displayed under the quiz title. In the intervention group, tailored persuasive explanations were appended below as quiz-feature-based explanations, matched according to the profile cluster. Students who saw these explanations were expected to be convinced of why the quiz was recommended and persuaded to solve it. For the convenience of the students, all quizzes used for recommendation range and their individual learning histories (○: correct, ×: incorrect, and ?: unsolved) were displayed as quiz stats below the

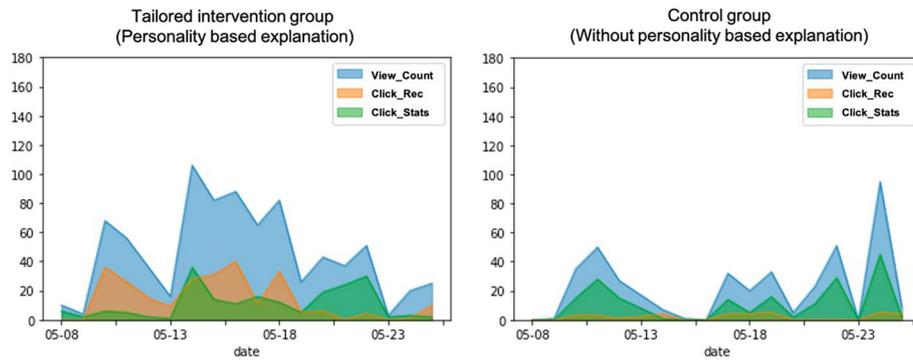


Fig. 6 Recommender usage for intervention and control groups

Table 4 System access and click-count statistics

Condition	Accessed	Clicked_Rec	Clicked Stats	CVR (%)
Tailored intervention group (personality-based explanation)	818	255	196	56.5
Control group (without personality-based explanation)	406	35	191	15.5

recommended quizzes. Students could also access the quizzes by clicking on the title of the quiz-stats list.

Results

Overview of recommender usage

As shown by the blue and red colors in Fig. 6, the number of accesses and clicks on the recommended questions were higher in the intervention group than in the control group. In the control group, the recommended quizzes were rarely used; instead, quiz lists were frequently used, indicated in green.

Table 4 shows the number of accessed systems (Accessed), number of clicks on the recommender system (Clicked-Rec), and number of clicked quizzes from the quiz-stats list (Clicked-Stats). The intervention group had approximately twice as frequent access as the control group and approximately seven times as many clicks on the recommended questions. We evaluated the effect of profile-based tailored explanation considering an indicator of effectiveness. When students solved a quiz, either from a recommendation or from the quiz-stats list, we defined the CVR (Zhang & Chen, 2020) as $\frac{\text{Clicked_Rec}}{\text{Clicked_Rec} + \text{Clicked_Stats}}$. As shown in Table 4, the CVR was 56.5% in the tailored intervention group (with profile-based explanation) and 15.5% in the control group (without profile-based explanation), approximately 3.65 times higher in the intervention group than in the control group. In our previous study of conventional recommendation with only an explanation using BKT parameters, the CVR was 6.17% for a summer-vacation homework experiment (Takami et al., 2022). A considerable improvement in the click rate was indicated for recommended quizzes with additional personality-based

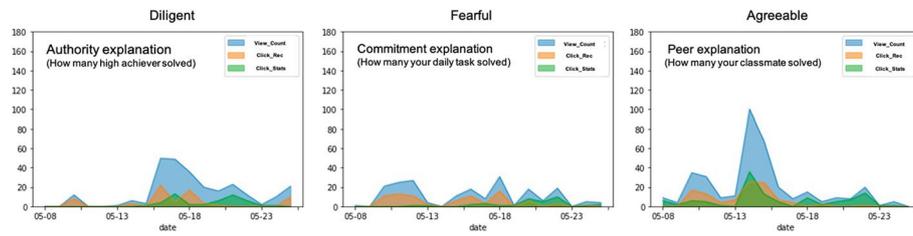


Fig. 7 Recommender usage for three intervention groups

Table 5 Access and click-count statistics for each profile segment

Condition		Accessed	Clicked_Rec	Clicked_Stats	CVR (%)
Tailored intervention group (personality-based explanation)	Diligent (N=32)	261	68	196	56.5
	Fearful (N=39)	199	82	35	70.1
	Agreeable (N=35)	358	105	113	48.2
Control group (without personality-based explanation)	Diligent (N=32)	137	29	55	34.5
	Fearful (N=39)	46	6	17	26.1
	Agreeable (N=35)	233	0	119	0.0

explanations, even considering whether the experimental period was the regular class period (this study) or the summer-vacation period (previous study).

Overview of recommender usage in each profile segment

We examined the use of the recommendation function for each cluster. Figure 7 shows the number of accesses and clicks on recommended questions and the status list for each profile cluster. There was some use of recommendation questions in all profile clusters, as indicated in red (Clicked_Rec). From the statistics of access and click counts in the tailored intervention group (Table 5, top), the Fearful group had the highest CVR (71.1%), compared to 48.2% for the Agreeable group and the overall CVR of 56.5%, as shown in Table 4. In the control group (Table 5, bottom), Agreeable students never used the recommended quizzes. Diligent and Fearful students had a CVR of approximately 30%, but the Clicked_Rec counts were fewer than in the tailored intervention group. These results indicate that tailored intervention was effective in increasing the overall number of clicks on recommendation questions, and that the intervention group had a higher number of clicks on recommendation questions than the control group.

Evaluation of individual recommended quiz usage

Thus far, we have considered the CVR by focusing on the total number of clicks to compare with our previous studies. We found higher recommended quiz counts than quiz lists. We now consider the number of times an individual recommended quiz was used, as shown in Fig. 8. In the box plots, the black lines within the box represent the median; the edges of the box represent the 25th and 75th percentiles of the data. Dots and whiskers display all data points and their ranges. The number of outlier dots differed significantly in the intervention group and the control group. The statistics for the intervention and control groups are summarized in Table 6. The intervention group had a mean of

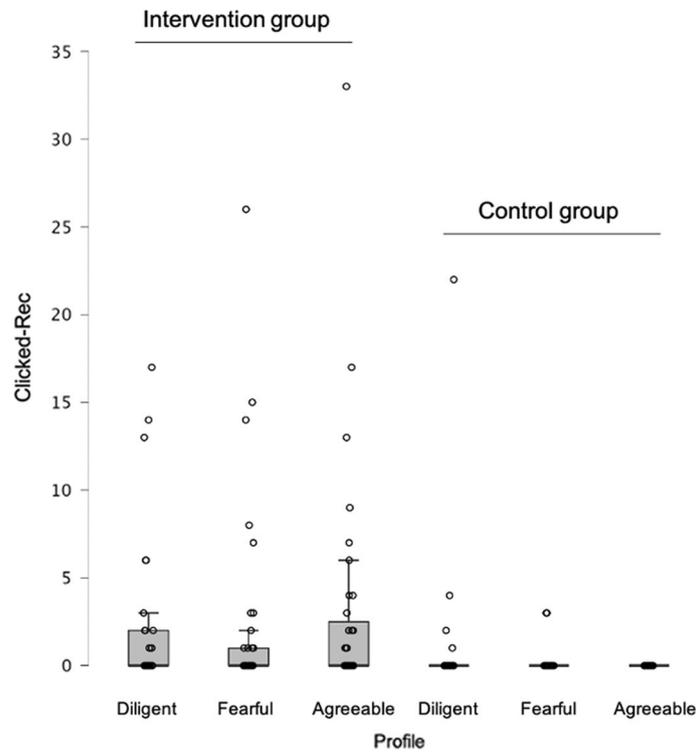


Fig. 8 Box plots of individual recommended quiz click-counts (Clicked_Rec) for each group

Table 6 Intervention and control group Clicked_Rec description

Condition	Mean	SD	SE
Tailored intervention group (N = 106) (personality-based explanation)	2.406***	5.470	0.531
Control group (N = 111) (without personality-based explanation)	0.315	0.315	0.205

Student's *t*-test, *df* = 215, *t* = 3.734, ****p* < 0.001

2.406 clicks, compared with 0.315 for the control group, a difference of approximately eight times (Table 6).

Evaluation of individual recommended quiz usage for each profile

We also evaluated each profile group (intervention and control groups). From Fig. 8, it is clear that some participants in all groups had zero clicks, with extreme values among the respondents. Thus, we conducted a Mann–Whitney *U* test, a non-parametric statistical test used to compare two independent groups, to determine if there was a statistically significant difference between them. The statistics for each profile group are summarized in Table 7. All intervention groups used the recommended questions > 2 on average. Considering that the average value in the control group was < 1 in all groups, we found that the intervention group had highly recommended quiz use for all profiles. In the intervention group, the Agreeable group had the highest mean recommended quiz click counts, whereas in the control group, the Agreeable group did not use the

Table 7 Each profile descriptive of individual recommended quizzes usage

Condition		Mean	Std	Max	Min
Tailored intervention group (personality-based explanation)	Diligent (N = 32)	2.125	4.420	17	0
	Fearful (N = 39)	4.420	5.290	26	0
	Agreeable (N = 35)	3.000	6.539	33	0
Control group (without personality-based explanation)	Diligent (N = 32)	0.644	3.325	22	0
	Fearful (N = 39)	0.182	0.727	3	0
	Agreeable (N = 35)	0	0	0	0

Table 8 Recommended quiz usage comparison within clusters

Mann–Whitney test	W	P	Rank-Biserial correlation
Diligent (intervention group–control group)	925.0	0.001*	0.285
Fearful (intervention group–control group)	802.5	0.005*	0.247
Agreeable (intervention group–control group)	–	–	–

The agreeable control group could not be tested because there were no clicks on the recommended quizzes

recommended quizzes. As shown in Table 5, the Agreeable group in the control group had 119 Clicked_Stats counts; students in this group did not use the recommended quizzes but used quizzes in the list (Fig. 5) on the recommendation page. As no recommended quizzes were used in this Agreeable group, we excluded it from a Mann-Whitney U test and we found significant differences between the intervention group and the control group (Table 8). These results suggest the effectiveness of personality-based tailored explanations in the educational recommender system.

Interaction pattern of each intervention group through process-mining

A more in-depth investigation was conducted using process-mining to clarify how the recommended quizzes were solved. DISC (Fluxicon, 2023) was used to identify prominent interaction processes for each of the three intervention groups. The process of interaction behaviors emerged through process-mining from interaction logs (Accessed, Clicked_Rec, ClickedStats, QUIZ_ANSER_CORRECT, and QUIZ_ANSER_WRONG), as shown in Fig. 9. Process-mining uses each logged interaction as a state, represented as a node in the graph, and a sequence (transition of one action to another) as the edge of the graph. The information in the node also provides the number of students performing a specific action. For example, 26 students in the Diligent intervention group accessed the recommender system (Fig. 9, top panel). Information on the edge represents the median time between the transition to the next action and the number of students with a specific transition pattern. For example, after clicking on the recommended quiz (Clicked_Rec in Fig. 9, top) and solving the quiz in a median time of 3.1 min, three students answered incorrectly (QUIZ_ANSWER_WRONG). Eleven students answered correctly (QUIZANSER_CORRECT), with a median time of 33 s.

Comparing the behavior of each group after the recommended question was clicked, as indicated by the red arrows, the median time to solve a question answered incorrectly

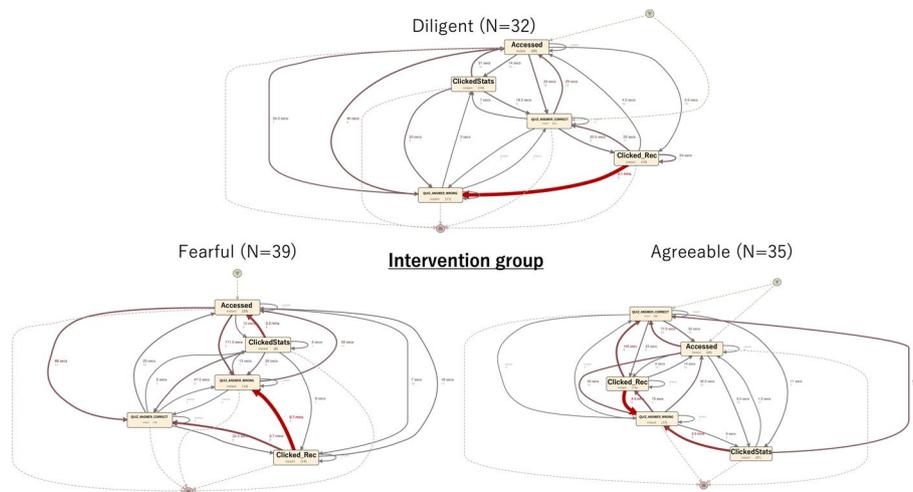


Fig. 9 Process-mining for each intervention group

was the highest for all groups, suggesting that more time was required to solve a difficult question. The time required to answer correctly was shorter for all groups. These results suggest that BKT comprehension estimation recommends questions of a moderate difficulty level, which is the likely reason why some questions were answered incorrectly and others were answered correctly. In the Fearful group (Fig. 9, bottom left), after ClickedStats were selected from a list of questions and the quiz opened, they returned to the original recommendation page (median 3.2 min for eight participants). This may be because they chose the questions themselves, wondered whether to answer them, and eventually returned to the original page without solving them. Such behavior may be characteristic of fearful traits and high anxiety tendencies.

Correlation of personality scale and recommended quiz usage in intervention group

We also examined the extent to which the personality scale was related to the number of clicks on the recommendation questions in the intervention clusters. No significant correlation was observed in the Diligent group. In the Fearful group, we found a significant negative correlation between Extraversion and Clicked-Rec, meaning that more extraverted fearful students tended to use the recommended quizzes less. This result suggests that for learners with high sociability and a tendency toward anxiety, commitment-type intervention may not be effective. Thus, it may be necessary to consider alternative explanations tailored to personality traits. In the Agreeable group, we found a positive correlation between Good_at_math and Rec_click. This result suggests that the Big Five personality scale and students' subjective perceptions of strengths or weaknesses in learning may be important in segmentation.

The correlational analyses suggest that categorization into three groups was somewhat coarse. To implement effective explanatory interventions, more detailed classifications and explanations tailored to specific personality traits must be considered. Additionally, in school settings, although group interventions based on personality are effective to a certain extent, individualized interventions may be necessary for learners who benefit less from group-based approaches.

Limitations

In this study, it was confirmed that adding explanations tailored to personality increased learner engagement. To determine whether each explanation was appropriate for each of the three groups, a comparison between matched and unmatched explanations was necessary because the study did not have a sufficiently large sample size to validate interventions that did not match the profile. A previous psychologically based tailoring study on public health found no significant differences between matched and unmatched intervention conditions (Hirai et al., 2016). Although further research is needed to validate the effectiveness of tailored interventions comparing matched and unmatched groups, this study indicated that compared with controls, additional matched tailored explanations according to personality were effective. In this study, although we used k-means clustering, there is room for verification as to whether similar clustering results can be achieved with different volumes and distributions of data, particularly in the context of class sizes ranging from approximately 200–500 students in a single grade level at one school.

Another limitation of the Big Five Inventory is that it is difficult to ask nearly 70 questions to students from K–12. Methods are required to reduce the burden of implementation, such as predicting personality traits from learning logs (Denden et al., 2018; Ghorbani & Montazer, 2015; Takami et al., 2022b). However, privacy concerns must be considered when implementing methods to estimate personal information. These issues should be treated as matters of privacy in decision-making processes (Acquisti & Grossklags, 2005; Kokolakis, 2017).

Discussion and conclusions

The main unique finding of this study was that personality-based tailored interventions aimed at increasing engagement with explainable recommender systems were significantly more effective than a conventional quiz character-based explanation-only approach in the A/B test condition (Table 6). The overall CVR of 56.5% (Table 5) in the intervention group was considerably higher than that reported in our previous study (CVR, 6.17). There are several possible explanations. In our previous study, the experiment was conducted in an environment where the teacher was not available during summer vacation; in this study, the experiment was conducted during a regular class period. In addition, in the previous study, the students were rushed to work on the assignment immediately before the end of summer vacation; thus, they may not have been able to fully utilize the recommended questions. Even discounting this, the intervention group CVR was much higher than that of the control group using the same means of explanation based on the quiz character.

Based on previous findings that conscientious people tend to be persuaded by people with authority (Alkış & Taşkaya Temizel, 2015), this study assumed that authority figures were the top performers. Authority is considered as a form of social influence; people tend to follow suggestions and recommendations from those with authority (Blass, 1991; Milgram & van Gasteren, 1974). Although top achievers may have authority, it is conceivable that others would also have authority, including teachers, school seniors with excellent grades, and those at the top of the hierarchy in school society. The Diligent

group providing authority explanation did not have as high a CVR (Table 5) or mean number of recommended quiz clicks (Table 7) as the other two groups. Thus, there is room for improvement as explanations from other authorities may be more effective.

The Fearful group had the largest overall CVR (Table 5); the average number of recommended quizzes for individual students was ranked second (Table 7). This means that some in the Fearful group used the recommended quizzes frequently and others did not. Heavy users of the recommended questions accounted for the number of times they used the recommended quizzes in this group. In addition, as shown in Fig. 10, the more extroverted fearful students tended to use fewer recommended quizzes, suggesting that more detailed tailoring interventions are needed for these students. Extraversion plays a crucial role in formation of social networks, primarily through what is termed as the 'popularity effect', suggesting that individuals with higher levels of extraversion tend to have a larger circle of friends than their introverted counterparts (Feiler & Kleinbaum, 2015). Thus, it may be effective to provide explanations to the Agreeable group, such as how many classmates solved the quizzes, to high-anxiety students with high extraversion.

The Agreeable group had the lowest CVR (48.2) (Table 5) in the tailored intervention condition, different from that in our previous report (6.17) (Takami et al., 2022). Recommendation questions were not used in the control group. This means that the peer-persuasive explanation (how many classmates solved the recommended quizzes) was effective for Agreeable students. In the Agreeable group, students who thought they were good at math used the recommended quizzes more often (Fig. 10, right). These results suggest that it is important to consider personality traits and skill level in persuasive explanations in education. In mathematics, math anxiety (Luttenberger et al., 2018) has become a major problem, and should be considered in development of persuasive educational systems.

Explanations of why an item is recommended are important in educational settings (Takami et al., 2022a). There are two main explanations: model-intrinsic and post hoc approaches (Zhang & Chen, 2020). In the model-intrinsic approach, the model mechanism is transparent, and the explanation indicates exactly how the model generates a recommendation interpretation from the recommender model algorithm. This approach is related to explainable AI (XAI) and has recently received attention in the field of education (Khosravi et al., 2022). In education, additional benefits of explanations from learning systems have been proposed (Ogata et al., 2024; Flanagan et al.,

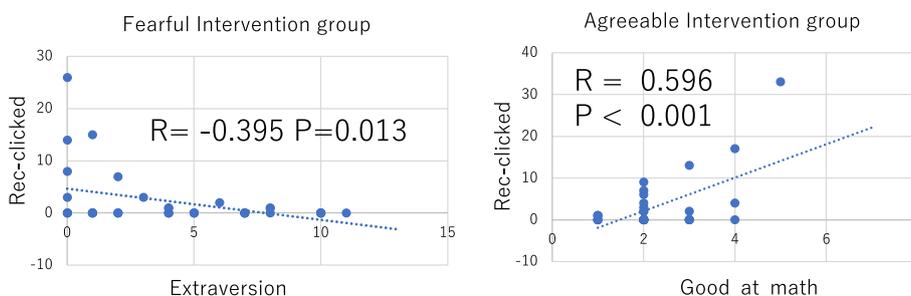


Fig. 10 Correlation between personality scales and Rec-clicked in intervention group

2021), such as encouraging student motivation to learn, leading to higher achievement. There are several possible reasons for use of explainable recommendation systems in education. A data-driven explanation based on learning history explains that a question is recommended based on mistakes made on it in the past. Using the knowledge model, this knowledge is related to other knowledge, and the quiz is recommended. Regarding persuasive explanations, we found that adding a tailored persuasive explanation to an explanation of the characteristics of the problem estimated from conventional learning history had a great effect on engagement. These results suggest that it may be effective to combine several explanation methods according to student characteristics such as personality, rather than using only one explanation method, to improve transparency, persuasiveness, and trustworthiness. New explanation methods tailored to personality can be considered as advancement of previous explainable recommendation research.

Our intervention was tailored using a personality trait-based approach based on responses to the Big Five psychological questionnaire. Previous learning analysis research (Matz et al., 2021) did not use the psychological personality trait scale, although some learning-style-related questionnaires were used for clustering. A design-based approach was used for university students; it was less robust than the A/B testing we used. The Big Five model of personality has been validated, although it has been argued that it does not capture the full range of human personalities, as it mostly concerns the more prosocial aspects of behavior (Paulhus & Williams, 2002), intelligence (Jensen, 1998), inhibition/activation (Carver & White, 1994), narcissism (Raskin & Hall, 1979), grit (Duckworth et al., 2007), and happiness (Lyubomirsky & Lepper, 1999). However, we did not examine these personality traits. Clustering using different personality measures and tailored interventions for each segment would be more effective for engagement. Tailored interventions including target segmentation for all aspects of learners may require improvement for future implementation. If empirical evidence confirms that certain types of explanations are more effective for specific learner profiles, development of a dataset that pairs learner types with effective explanations is a logical avenue for research. Such a dataset could be instrumental in fine-tuning large-scale language models such as ChatGPT (OpenAI, 2022) and Llama 2 (Meta, 2023), enabling them to generate highly accurate and trustworthy explanations tailored to individual learner types.

A personality-based, segmented tailored intervention for students designed to increase engagement with explainable recommender systems was significantly more effective than conventional explanations. These results suggest that personality-based explanations in the recommender approach are effective for e-learning engagement and imply improved trustworthiness in the AI learning system.

Acknowledgements

We would like to thank the high school students and teachers who participated in this study. The data of this paper and preliminary analysis has been presented in Takami et al. (2023).

Author contributions

KT, BF, and HO contributed to the research conceptualization and methodology. KT and YD conducted experiments and collected the data. KT analyzed data and wrote the first draft. YD, BF and provided comments to improve the manuscript. All authors read and approved the final manuscript.

Funding

This work was partly supported by JSPS Grant-in-Aid for Early-Career Scientists JP23K17012, JSPS Grant-in-Aid for Scientific Research (B) JP23H01001, JP22H03902, JP20H01722, JSPS Grant-in-Aid for Scientific Research (Exploratory) JP21K19824, and NEDO JPNP20006.

Availability of data and materials

Not applicable. The data of this study is not open to the public due to participant privacy.

Declarations**Competing interests**

The author declares no competing interests.

Received: 30 June 2023 Accepted: 21 November 2023

Published online: 11 December 2023

References

- Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1), 26–33.
- Alkış, N., & Taşkaya Temizel, T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences*, 87, 147–152.
- Asendorpf, J. B. (2002). The puzzle of personality types. *European Journal of Personality*, 16(1_suppl), S1–S5.
- Badrinath, A., Wang, F., & Pardos, Z. (2021). *pyBKT: An accessible python library of Bayesian knowledge tracing models* (pp. 468–474).
- Barria-Pineda, J., Akhuseyinoglu, K., Želem-Čelap, S., Brusilovsky, P., Milicevic, A. K., & Ivanovic, M. (2021). Explainable recommendations in a personalized programming practice system. *Artificial Intelligence in Education*, 64–76.
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, 60(3), 398–413.
- Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1998). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences*, 26(1), 129–140.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, 67(2), 319–333.
- Cialdini, R. B. (2001). The science of persuasion. *Scientific American*, 284(2), 76–81.
- Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298, 103503.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278.
- Dai, Y., Flanagan, B., Takami, K., & Ogata, H. (2022). Design of a User-Interpretable Math Quiz Recommender System for Japanese High School Students. In *Proceedings of the 4th workshop on predicting performance based on the analysis of reading behavior*.
- Dai, Y., Takami, K., Flanagan, B., & Ogata, H. (2023). Beyond recommendation acceptance: Explanation's learning effects in a math recommender system. *Research and Practice in Technology Enhanced Learning*, 19, 020.
- Denden, M., Tlili, A., Essalmi, F., & Jemni, M. (2018). Implicit modeling of learners' personalities in a game-based learning environment using their gaming behaviors. *Smart Learning Environments*, 5(1), 1–19. <https://doi.org/10.1186/s40561-018-0078-6>
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.
- Feiler, D. C., & Kleinbaum, A. M. (2015). Popularity, similarity, and the network extraversion bias. *Psychological Science*, 26(5), 593–603.
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal*, 10(4), 469–484.
- Flanagan, B., Takami, K., Takii, K., DAla, Y., Majumdar, R., & Ogata, H. (2021). *EXAIT: A symbiotic explanation learning system* (pp. 404–409).
- Fluxicon. (2023). *DISCO[Computer software]*. Retrieved 2023, from <https://fluxicon.com/disco/>.
- Fogg, B. J. (2002). Persuasive technology: Using computers to change what we think and do. *Ubiquity*, 2002(December), 2.
- Ghorbani, F., & Montazer, G. A. (2015). E-learners' personality identifying using their network behaviors. *Computers in Human Behavior*, 51, 42–52.
- Giluk, T. L., & Postlethwaite, B. E. (2015). Big five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences*, 72, 59–67.
- Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education*, 20(3), 261–273.
- Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105–120.
- Hirai, K., Ishikawa, Y., Fukuyoshi, J., Yonekura, A., Harada, K., Shibuya, D., Yamamoto, S., Mizota, Y., Hamashima, C., & Saito, H. (2016). Tailored message interventions versus typical messages for increasing participation in colorectal cancer screening among a non-adherent population: A randomized controlled trial. *BMC Public Health*, 16, 431.
- Jensen, A. R. (1998). *The factor*. Westport, CT: Praeger.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/t07550-000>
- 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.
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122–134.

- Kreuter, M. W., Strecher, V. J., & Glassman, B. (1999). One size does not fit all: The case for tailoring print materials. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, 21(4), 276–283.
- Lopez, F. G., & Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *The Career Development Quarterly*, 41(1), 3–12.
- Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management*, 11, 311–322.
- Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, 46(2), 137–155.
- Matz, R., Schulz, K., Hanley, E., Derry, H., Hayward, B., Koester, B., Hayward, C., & McKay, T. (2021). Analyzing the efficacy of ECoach in supporting gateway course success through tailored support. In *LAK21: 11th international learning analytics and knowledge conference* (pp. 216–225).
- Meta. (2023). *Llama 2*. <https://huggingface.co/meta-llama/Llama-2-70b-chat-hf>
- Milgram, S., & van Gasteren, L. (1974). *Das Milgram-experiment*. Hamburg: Rowohlt.
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3), 393–444.
- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43(5), 971–990.
- Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., & Takii, K. (2024). EXAIT: Educational eXplainable Artificial Intelligent Tools for personalized learning. *Research and Practice in Technology Enhanced Learning*, 19, 019.
- OpenAI. (2022). *Introducing ChatGPT*. <https://openai.com/blog/chatgpt/>
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563.
- Rahdari, B., Brusilovsky, P., & Thaker, K. (2020). Using knowledge graph for explainable recommendation of external content in electronic textbooks. *ITextbooks*.
- Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, 45(2), 590.
- Rimer, B. K., Conaway, M., Lyna, P., Glassman, B., Yarnall, K. S. H., Lipkus, I., & Barber, L. T. (1999). The impact of tailored interventions on a community health center population. *Patient Education and Counseling*, 37(2), 125–140.
- Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: Three replicable personality types. *Journal of Personality and Social Psychology*, 70(1), 157–171.
- Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2), 47–53.
- Sohl, S. J., & Moyer, A. (2007). Tailored interventions to promote mammography screening: A meta-analytic review. *Preventive Medicine*, 45(4), 252–261.
- Takami, K., Flanagan, B., Dai, Y., & Ogata, H. (2021). Toward educational explainable recommender system: explanation generation based on Bayesian knowledge tracing parameters. In *29th International Conference on Computers in Education Conference Proceedings* (Vol. 2, pp. 532–537).
- Takami, K., Dai, Y., Flanagan, B., & Ogata, H. (2022). Educational explainable recommender usage and its effectiveness in high school summer vacation assignment. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 458–464).
- Takami, K., Flanagan, B., Majumdar, R., & Ogata, H. (2022). Preliminary Personal Trait Prediction from High School Summer Vacation e-learning Behavior. In *Proceedings of the 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior*.
- Takami, K., Flanagan, B., Dai, Y., & Ogata, H. (2022). Toward trustworthy explainable recommendation: personality based tailored explanation for improving e-learning engagements and motivation to learn. In *Companion Proceedings 13th International Conference on Learning Analytics & Knowledge (LAK23)*.
- Tempelaar, D., Rienties, B., & Nguyen, Q. (2021). Dispositional learning analytics for supporting individualized learning feedback. In *Frontiers in education*, vol. 6. <https://doi.org/10.3389/educ.2021.703773>
- Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 353–382). Springer.
- Tsiakas, K., Barakova, E., Khan, J. V., & Markopoulos, P. (2020). BrainHood: towards an explainable recommendation system for self-regulated cognitive training in children. In *Proceedings of the 13th ACM international conference on Pervasive technologies related to assistive environments*, 1–6.
- Wall, H. J., Campbell, C. C., Kaye, L. K., Levy, A., & Bhullar, N. (2019). Personality profiles and persuasion: An exploratory study investigating the role of the Big-5, Type D personality and the Dark Triad on susceptibility to persuasion. *Personality and Individual Differences*, 139, 69–76.
- Yu, R., Pardos, Z., Chau, H., & Brusilovsky, P. (2021). Orienting students to course recommendations using three types of explanation. In *Adjunct proceedings of the 29th ACM conference on user modeling, adaptation and personalization* (pp. 238–245). Association for Computing Machinery.
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends in Information Retrieval*, 14(1), 1–101.

Publisher's Note

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