

AI-BASED QUIZ SYSTEM FOR PERSONALISED LEARNING

X. Wang¹, S. Wrede², L. van Rijn², J. Wöhrle²

¹German Research Center for Artificial Intelligence (DFKI) (GERMANY)

²Center of Advanced Technology for Assisted Learning and Predictive Analytics (GERMANY)

Abstract

This paper has presented an AI-assisted quiz system, *iQS*, with the capability of creating individual quizzes and providing instant personal feedback, to serve self-regulated learning and distant learning. *iQS* is designed to be easily adopted by different learning management systems (LMS, such as Moodle or MOOCs). Its core algorithms can be reused directly, only specific domain knowledge and learners' learning data need to be provided to run in a new specific domain. As a research instance, *iQS* has been developed and tested inside the Moodle learning platform of a university and positive feedback has been received from students and teachers.

Keywords: Quiz system, knowledge modelling, Ontology, personalised quizzes, self-regulated learning.

1 INTRODUCTION

For the current digitalized online learning, personalized learning ([6], [19]) and self-regulated learning (SRL) are both very active research topics. It is said that the most effective learners are self-regulating [7]. Usually, a self-regulated learner normally monitors, directs, and regulates actions toward goals of information acquisition, expanding expertise, and self-improvement [8]. Feedback has a powerful influence on SRL processes and affects students' cognitive engagement with tasks ([2], [12]). Our research overall is committed to studying individual learning differences in different learning stages and how to provide the most adaptive and accurate feedback in real-time to improve students' learning and SRL skills.

One of our studies focuses on how to apply the latest AI technologies in higher education via practical applications. In this paper, our scenario is that an intelligent AI-based quiz system is required by university students, who are basically enrolled in distance learning courses using blended learning. Over half of them study part-time, have jobs or families, learn for a better career or certain interests, and are mainly learning by themselves. AI-based online learning systems have proven to be powerful tools for facilitating self-regulated learning processes ([14], [16]).

In the initial phase of our research, we found out that the learning management system (LMS) currently in use only records learning progress digitally in a coarse-grained way; allows digitized learning content merely on a small scale; and has barely any personalized setting for students; and has few feedback directly from tutors. Therefore, the data sets to be collected for the purpose of data analysis are generally of poor quality, which is incomplete and fragmented or even missing. In addition, there are only a small number of quizzes that are manually created by tutors from textbooks and a fixed feedback made available to all students without distinction.

After reviewing the current states of assessment/quiz systems (9), [13], [22]), including some commercial ones (i.e., *Questgen*, *Quillionz*), we couldn't find an AI-based quiz system, which is ready to use and easy to expand with new AI capabilities. Considering our specific requirements for personalized learning and knowledge self-assessment, we proposed a scalable AI-powered quiz system, *iQS*. Initially, the following research questions are stated:

- Q1: What does a general, scalable, and AI-based quiz system look like?
- Q2: How does it provide personalized and intelligent feedback in real time?
- Q3: What is the learning experience of students using it for their self-regulated learning? and are they motivated to engage more with learning?

This study tries to answer the above-stated research questions but is not limited to them. Preliminary results of our research have been published [18], which have covered the quiz generation algorithm and knowledge feedback algorithms. It presented *iQS* only from the perspective of students. However, this paper presents the whole AI-based quiz system and explains its innovation from the teaching point of view, which is, how a tutor creates quiz items, how much effort is involved, and which support the tutor can get.

2 METHODOLOGY

2.1 Domain Knowledge and Quiz Modelling

iQS has built on top of two basic significant mechanisms, namely, Domain Knowledge (DK) and an innovative Quiz Ontology (QO). Obviously, well-defined domain knowledge with rich semantics on common prior or implicit intuitions is the key for AI to be applied in many application domains ([1], [20]).

In our research, we had more than 3 tutors as domain experts who went through at least four courses including their auxiliary textbooks, and manually built four ontologies respectively. They have not only defined the key knowledge concepts with very fine granularity but also linked them to concrete learning material in various learning formats (i.e., text, audio, videos, slides, and images). Such domain knowledge lays the foundation for our work, enabling us to track and precisely locate students' knowledge errors, and to calculate their knowledge mastery distribution (i.e., what knowledge a student has learned and how well he/she masters it, see Fig. 3). It is exactly why we can provide intelligent knowledge-based recommendations for a specific student's current learning.

The second cornerstone is our Quiz Ontology, where our innovation lies. We have summarized the following innovations: (1) a quiz question consists of two parts, one question (aka, question body), and a set of options. The separation of the question body and its options makes it possible to dynamically form personalized questions in an extremely flexible way. One option can loosely or semantically link to multiple questions and vice versa. Questions and options are saved in different data pools, respectively. (2) A quiz question will remain abstract until a set of concrete options is to be formed on the fly. (3) Each option is required to semantically link to one or multiple knowledge concepts and learning materials, i.e., only when all its options are specified, a question will be clear about what concrete knowledge it is testing. (4) a quiz consists of a set of quiz questions and will be dynamically generated on the fly for the individual students.

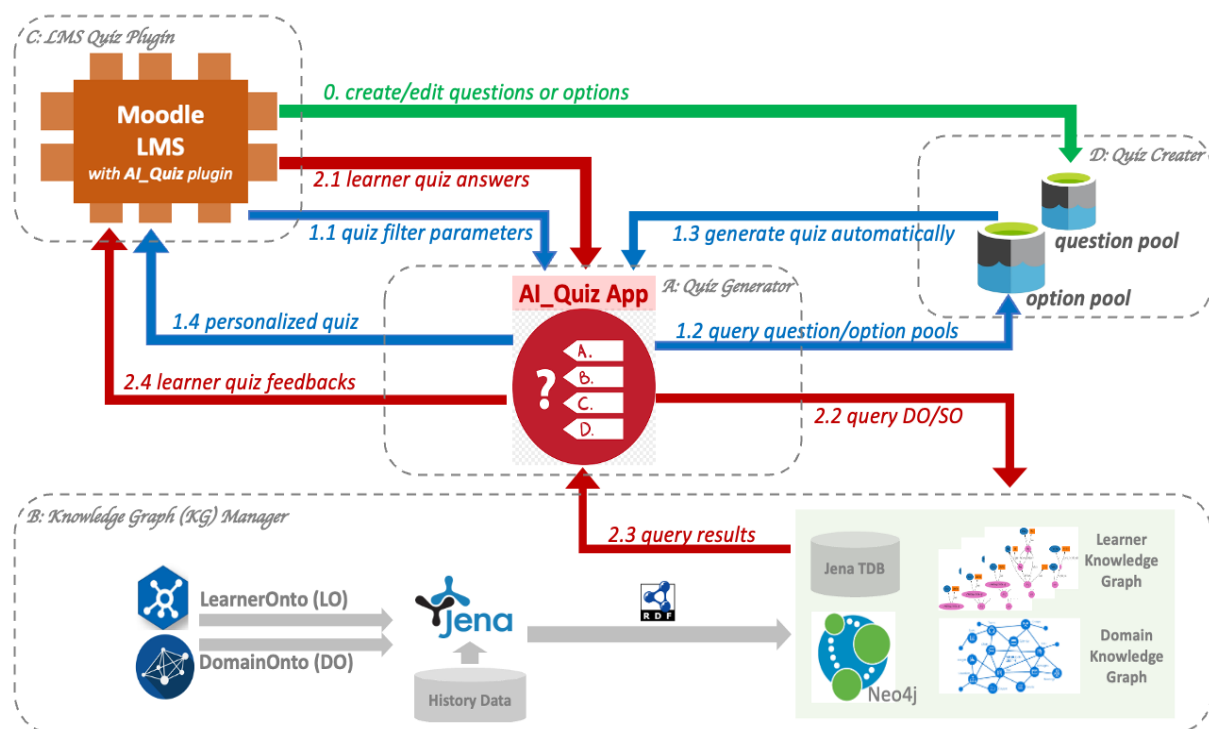


Figure 1. The AI-based quiz system (iQS).

In summary, designing quiz questions in this way is to guarantee the following: (1) different students get different quizzes, (2) the same student gets the same question body but with different options at different times, and (3) any quiz is generated fully based on students' own knowledge mastery levels at that very moment. (See proof in [18]).

2.2 The AI-based Quiz System

First, we briefly overview the *i*QS and the relationship among its four components, and then we address each of them in detail. The overall architecture of *i*QS consists of four major components:

- **Component A:** the quiz generator (QG) adaptively generates individual quizzes on the fly based on a set of real-time students' personal quiz filters (such as constraining the knowledge coverage to be tested, difficulty levels, competence levels, and so on). Besides, it provides knowledge-based feedback or recommendations instantly to students' answers (see Section 2.2.1).
- **Component B:** the knowledge graph manager (KGM) is not only in charge of the domain knowledge (DM) of courses but also manages learners' personal knowledge graphs (*p*KG) via tracking students' learning processes and results (see Section 2.2.2).
- **Component C:** the LMS quiz plug-in (QP) enables *i*QS to be seamlessly embedded into concrete learning management systems (see Section 2.2.3).
- **Component D:** the quiz content creator (QCC) is responsible for semi-automatically creating quiz question building blocks, i.e., quiz question bodies, and question options (see Section 2.2.4). Tutors can use this component manually to create new questions or only separate options or to edit the existing ones from the either question pool or option pool.

Fig. 1 illustrates the working process from the student's point of view (see the blue lines): (1.1) students can take the system-suggested filter parameters as default, or they are free to set their own to meet their quiz purposes (e.g., targeting on specific knowledge areas or the topics where they have done wrong before). So far 10 parameters have been considered, see the Fig. 2; (1.2) taking students' filters as input, sending queries to the question pool and option pool, and applying the quiz question generation algorithm (defined in [18]), concrete quiz questions are then to be formed; (1.3) quiz questions are generated, either all at once, either one at a time (with adaptive difficulty); (1.4) these quizzes are offered to students inside the quiz plug-in.

Figure 2. The personal quiz filter with 10 parameters.

The underlying core idea is, via using *i*QS, at any moment we can understand who exactly learns what or does not, and at what competence mastery level. Therefore, the more *i*QS is used, the more accurate and intelligent the default filter becomes, since *i*QS is constantly tracking and analyzing students' learning and learning results. In future research, we can go further to predict, e.g., when and who does or does not do what activities to achieve certain mastery levels. As mentioned above, the first three of the components considered in the sequel were developed previously [12]. We summarize them briefly here to ensure the completeness of this paper and try to view them from a different angle. As said, the first two components

as a preliminary result have been developed and published. To ensure the completeness of this paper, we summarize them briefly here, while trying to present them from the angle of teaching.

2.2.1 The Quiz Generator (QG)

The innovation of QG lies in its two core algorithms, one for generating personal quizzes (namely, PQ) and the other for providing customized knowledge-based feedback for students' current learning or recommendations for their next step of learning (namely, KR). Specifically, to generate a quiz for a student i ($1 \leq i \leq k$), the PQ algorithm requires the following inputs:

- a vector of 10 filter parameters being either quiz or knowledge related,
- a matrix of options, with n options from the option pool and each option having m linked knowledge concepts,
- a vector of knowledge concepts, with m knowledge concepts from the domain knowledge and each having a difficulty level value,
- the student's i learning cube over quizzes (SQ), where the entry SQ_{ij} tracks his/her latest three learning records for quiz option j ($j \leq n$); and each learning record capturing the information on whether options were answered successfully and the time required for answering,
- a vector of the student's knowledge mastery values, tracking the student's mastery levels of each knowledge concept.

Basically, with the given filter parameters, after a series of data normalizations, transformations, and matrix multiplication, the final winner options will stand out [18]. Due to the semantic links between question bodies and options, the candidate questions are substantively assembled as well eventually.



Figure 3. Example of a learner's personal knowledge graph (pKG) (partially excerpted)

On the other hand, the KR algorithm is fully based on a set of rules to generate knowledge feedback. It distinguishes between the two cases of whether an answer received is correct or incorrect. When a student submits an incorrect answer, the feedback will focus on how to correct and improve his/her current knowledge. Conversely, when his/her answer is correct, then the system will recommend new further knowledge for him/her to learn from there. Moreover, since each option semantically links to multiple knowledge concepts, by cross-checking the student's answer options, it is possible for us to accurately detect the exact position of the student's knowledge errors if any.

More importantly, tests with two simulated data sets showed that the algorithms are quite efficient, as it takes less than 2 seconds to generate a quiz with 30 questions for a student on both the small data set and the big one (as published). Where, the simulated big data set contains some 300 key knowledge concepts, 10K students, and 100K options; the ratio of one option to belong to multiple questions is temporarily kept around 10% ~ 15%.

2.2.2 The Knowledge Graph Manager (KGM)

Knowledge tracking is crucial for inferring the skill mastery of students and predicting their performance to adjust the learning accordingly. Since knowledge concepts and difficulty levels have been specified inside questions, individual knowledge mastery able to be tracked in our system, which conforms to the requirements of the interpretable knowledge tracking model proposed by [15] as well.

The KGM component is used to manage the domain knowledge and be responsible for creating and maintaining the students' individual knowledge graphs. For illustration purposes only, for example, after a while of self-regulated learning, the personal knowledge graph of student i might be as in Fig. 3. The green nodes represent the knowledge that he/she has learned, and the brown ones represent what have not learned yet. The size of the nodes is used to distinguish mandatory from optional knowledge. Each node is also associated with a knowledge mastery value of this student; for example, 0.80 means that the mastery level 80% is currently achieved on this knowledge concept. In addition, a small summary report is provided to show the student's personal progress. For example, compared to his/her previous quiz history, the student's knowledge mastery increased by +3.0%, even though the competency level did not change after the current quiz.

2.2.3 The Quiz Plug-in (QP)

*i*QS is designed for general purposes and versatile domains from the very beginning. Particularly for the current environment of micro-services and containers as mainstream application architectures, we fully take into account the scalability, flexibility, and loose coupling of our micro-services. Therefore, our components are developed as plug-in micro-service for current popular LMS, such as Moodle1 or MOOCs.

Since our current test bed is a Moodle learning system, a research instance of *i*QS has been implemented as a Moodle plug-in. Preliminary assessment results are to be presented in Section 3.

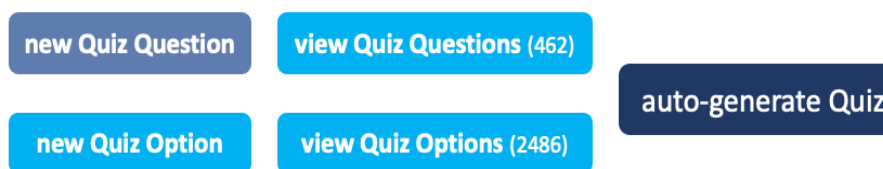


Figure 4. Five modules of the quiz content creation component

2.2.4 The Quiz Content Creator (QCC)

As the focus of this paper, the quiz content creator component represents our innovations in terms of teaching. It is completely different from traditional quiz creation, where a tutor or instructor creates a single quiz for all students with the same general feedback. Such a quiz cannot be adaptive to all students' learning states, nor can it dynamically interact in real time. Our AI-based QCC aims not only to free tutors from heavy manual work but also from impossible tasks (e.g., big mounts of real-time personal quizzes with intelligent feedback), allowing them to pay more attention to the design of teaching content itself.

Figure 5. Quiz content creation editor: (left) example of creating a new question and managing its options, (right) example of creating a new option and manually linking to a question.

The QCC contains five functional modules (see Fig. 4). The modules of new quiz question and new quiz option are used to create questions and options, respectively; the modules view quiz question and view quiz option allow tutors to view/edit existing questions or options. For example, there are 2486 options right now in the pool. The last module auto-generate quiz is used to automatically generate quiz options from domain knowledge and learning resources, for example, from textbooks.

2.3 Evolution of Quiz Content Creation

To boot up *iQS*, certain prepared quiz questions need to be created manually initially. Meanwhile, these data can also be used as a golden standard of later analytics. Therefore, the QCC component will go through two phases, one involving manual work and then a fully automatic one.

2.3.1 Instructor-involved Quiz Question Creation

Unlike existing quiz systems, multiple tutors can operate *iQS* at the same time, and the created questions and options are directly saved to the question and option data pools, respectively. Currently, it supports quite many types of questions such as Multiple Choice, Single Choice, True/False, Matching/ Ordering, Image Drag/Drop, or Gap Select.

Although a complete question (including a single question body and several options) can be created here, only the part of the question body is mandatory. As shown in Fig. 5(a), a tutor T only needs to set the type of question, the score, the title, and the question body. The remaining fields will mostly be auto-filled once some options have been linked to the specific question. For example, the fields of course ID(s), knowledge concepts, competence level, and difficulty level, will be filled with either the sums or the maximum values of all the linked question options.

The three ways to link an option to the current question body are: (1) to search for one in the option pool; (2) to create a completely new one; (3) to browse the system-suggested ones by clicking on the number next to the button (auto-suggested) correct options. Last but not least, a tutor also can provide a short text as content of further recommendation in case the question was answered correctly. Such short texts are to be collected and later used to fine-tune our NLP-based feedback generation algorithm.

Similarly, there are only a few inputs required when creating an option. For instance, in Fig. 5(b), a tutor can independently create a question option by defining the option body, the knowledge concepts to be tested, the difficulty level, the competence level, and short texts as feedback to correct and/or incorrect answers. At this stage, the tutor also has the opportunity to manually link the option to existing questions

using the search function. Notably, when an option is assigned to a specific question body, two pieces of information must be given, that is the correct answer to this question and the score weight.

2.3.2 Automatic Quiz Creation

Automatically creating question bodies and options for the two data pools from raw course materials will be our next step. Since automatically generating quizzes is already a well-researched topic, quite some work was done in this area by fusing technologies of neural networks (NNs) and natural language processing (NLP) ([4], [5], [9], [17], [21]). For example, [17] uses NLP and optical character recognition (OCR) to extract keywords from uploaded text images (e.g., scanned books) and from the internet to generate facts-based multiple-choice questions (MCQ). Moreover, [4], [5] and [21] worked on generating distractors of MCQ from free text with different methods, such as a point-wise ranking support vector machine, a list-wise ranking neural network, and a ranking generative adversarial network (GAN). Their experiments also got good results. We may consider directly applying similar approaches in our context next step.

3 USER EXPERIENCE EVALUATION RESULTS

As a Moodle plugin, *iQS* has undergone multiple progressive iterations, which have been thoroughly tested by university students. Specifically, *iQS* was first released to students in the winter semester of 2021/2022 for a duration of 17 days and 111 quizzes were created by 69 students eventually. Two surveys were carried out by a group of tutors and students from an educational science course alongside several subject experts. The evaluation process has been presented in Fig. 6 and every evaluation has a different target group with different focuses using different measurements.

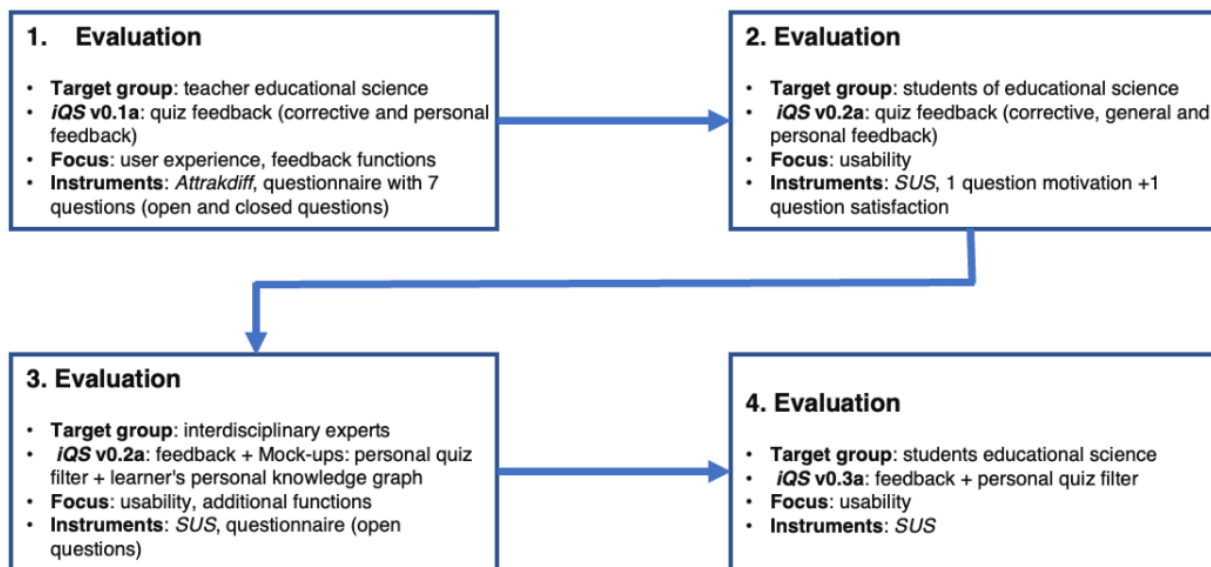


Figure 6. Evaluation process of *iQS*.

Basically, the system usability scale (SUS) [10] as a reliable tool has been chosen for measuring the usability of *iQS*. The SUS is made up from 10 items, which are cumulated for a final score. Furthermore, single item measures assessed students' satisfaction and their perceived impact on motivation when engaging with recommended learning materials on a 5-point Likert scale ranging from 0 (very low) to 4 (very high). For the first survey, the *iQS* v0.2a was rated by 35 participants with high usability ($M = 76.28$, $sd = 15.64$), corresponding to a grade of B, which is quite good already (see Table. 1). Additionally, they assessed the *iQS* v0.2a as motivating, to engage with the learning material ($M = 2.94$, $sd = 0.80$) and were satisfied with the current state of the *iQS* v0.2a ($M = 3.09$, $sd = 0.51$) (see Table. 2).

Table 1. SUS scores by iQS versions.

	iQS v0.2a (n=35)		iQS v0.3a (n=46)	
	<i>M</i>	<i>sd</i>	<i>M</i>	<i>sd</i>
Students	76.28	15.64	78.10	12.07
Experts	56.59	20.90		

In the second survey, by then iQS v0.3a has integrated more functional modules, including visualizing a learner's personal knowledge graph and the quiz filter with more parameters that students can customize their individual quizzes. Meanwhile, we continuously received the usability rate from our 11 participating experts, which is relatively low ($M = 56.59$, $sd = 20.90$), resulting in a grade of D. Based on experience, scores below C (68 points or lower) mostly referred to problems understanding the presented concepts and constructs in the mock-ups. It might explain why the experts were more critical of the usability of the iQS than students (Cohen's $d = -1.07$).

The fourth evaluation in the second survey took place in the summer semester of 2022. 771 students were offered to access the iQS for a period of 39 days. In that timeframe, 46 students generated 119 quizzes in all and answered the questionnaire. Again, we got in a high usability rate ($M = 78.10$, $sd = 12.07$) equivalent to a grade of B+. Compared to the previous assessment, it shows that iQS is motivating students to engage more in studying the recommended learning material ($M = 3, 13$, $sd = 0, 78$), and students were more satisfied with the tool ($M = 2.83$, $sd = 0.77$) as well.

Table 2. Students' ratings on motivation and satisfaction.

	iQS v0.2a (n=35)		iQS v0.3a (n=46)	
	<i>M</i>	<i>sd</i>	<i>M</i>	<i>sd</i>
Motivation	2.94	0.80	3.13	0.78
Satisfaction	3.09	0.51	2.83	0.77

So far, three versions of iQS (v0.1a~0.3a) were evaluated based on their initial usage. Although we were prepared for some criticism due to the incompleteness of functionality in the initial versions, the results are rather positive. To reflect specifically on the more critical feedback we received, influencing factors are possible: (1) the system at this stage focuses on the provision of functionality rather than high usability; (2) additional information on the system's usage (e.g., more communication between users and system designer, user documentation, tutorials) has not yet been introduced.

4 CONCLUSIONS

A novel AI-based quiz system has been presented in detail in this paper, which aims to support students' self-regulated learning with completely personalized quizzes. It has been designed to be a pervasive, scalable, and AI-based knowledge assessment system and potentially to be used in various application domains. As the first implementation, the iQS system has been initially tested and evaluated by university students and tutors rendering a first batch of positive objective feedback. Students showed certain satisfaction and stated that iQS could motivate them to engage more with the learning material than before.

ACKNOWLEDGEMENTS

This research is part of the AI.EDU Research Lab project led by Prof. Claudia de Witt and Prof. Niels Pinkwart and founded by the Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA), at the FernUniversität in Hagen.

REFERENCES

- [1] A. Agarwal, D. S. Mishra, S. V. Kolekar, "Knowledge-based recommendation system using semantic web rules based on learning styles for MOOCs", *Cogent Engineering*, vol. 9, 2022568, 2022. doi:10.1080/23311916.2021.2022568.
- [2] B. Wisniewski, K. Zierer, J. Hattie, "The power of feedback revisited: A meta-analysis of educational feedback research", *Frontiers in Psychology*, vol. 10, 2020.
- [3] C. Jouault, K. Seta, Y. Hayashi, "Content-dependent question generation using lod for history learning in open learning space", *New Generation Computing*, vol. 34, pp. 367–394, 2016. doi:10.1007/s00354- 016- 0404- x.
- [4] C. Liang, X. Yang, N. Dave, D. C. Wham, B. Pursel, C. L. Giles, "Distractor generation for multiple choice questions using learning to rank", in: BEA@NAACL-HLT, 2018.
- [5] C. Nwafor, I. Onyenwe, "An automated multiple-choice question generation using natural language processing techniques", *Int. J. on Natural Language Computing*, vol. 10, pp. 1–10, 2021. doi:10.5121/ijnlc.2021.10201.
- [6] C. Walkington, M. Bernacki, "Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions", *J. of Research on Technology in Education*, vol. 52, pp. 235–252, 2020. doi:10.1080/15391523.2020.1747757.
- [7] D.L. Butler, P.H. Winne, "Feedback and self-regulated learning: A theoretical synthesis", *Review of Educational Research*, vol. 65, pp. 245–281, 2018. doi:10.3102/00346543065003245.
- [8] E. Panadero, "A review of self-regulated learning: Six models and four directions for research", *Front Psychol.* 2017, vol. 8, 2017. doi:10.3389/fpsyg.2017.00422.
- [9] G. Kurdi, J. Leo, B. Parsia, U. Sattler, S. Al-Emari, "A systematic review of automatic question generation for educational purposes", *Int. J. of Artificial Intelligence in Education*, vol. 30, pp. 121–204, 2020. doi:10.1007/s40593- 019- 00186-y.
- [10] J. Brooke, "SUS: A quick and dirty usability scale", in: P. W. Jordan, B. Thomas, B. A. Weerdmeester, I. McClelland (Eds.), *Usability evaluation in industry*, CRC Press, London, pp. 189–194, 1996.
- [11] K. Vachev, M. Hardalov, G. Karadzhov, G. Georgiev, I. Koychev, P. Nakov, "Generating answer candidates for quizzes and answer-aware question generators", *CoRR abs/2108.12898*, 2021.
- [12] L.A. Lim, D. Gasevic, W. Matcha, N. Ahmad Uzir, S. Dawson, "Impact of learning analytics feedback on self-regulated learning: Triangulating behavioural logs with students' recall", 11th Int. Learning Analytics and Knowledge Conf. (LAK21), Association for Computing Machinery, New York, pp. 364–374, 2021. doi:10.1145/3448139.3448174
- [13] M. Elshafey, T. Ghoniemy, "The development of QMMS: A case study for reliable online quiz maker and management system", *Future Computing, and Informatics J.*, vol. 6, pp. 82–96, 2021. doi: 10.54623/fue.fcij.6.2.3.
- [14] R. Luckin, "Towards artificial intelligence-based assessment systems", *Nature Human Behaviour*, vol. 1, 2017. doi:10.1038/s41562- 016- 0028.
- [15] S. Minn, J. Vie, K. Takeuchi, H. Kashima, F. Zhu, "Interpretable knowledge tracing: Simple and efficient student modeling with causal relations", *CoRR 2021. arxiv.org/abs/ 2112.11209*.
- [16] S.M. Lorenzo, A.-A. Antonio, G.-H. Laura, "Analysis of online quizzes as a teaching and assessment tool", *J. of Technology and Science Education*, vol. 2, pp. 39–45, (2012). doi:10.3926/jotse.30.
- [17] V. Gupta, H. Pant, A. Chaurasia, A. Dwivedi, S. Singh, P. Singh, "Automatic quiz generator", *International Journal of Research in Engineering*, *Science and Management*, vol. 4, pp. 54–56, 2021.
- [18] X. Wang, H. Li, A. Zimmermann, N. Pinkwart, S. Wrede, L. Van Rijn, C. De Witt, B. Baudach, "IFSE - personalized quiz generator and intelligent knowledge recommendation", 2022 IEEE 16th Int. Conf. on Semantic Computing (ICSC), pp. 201–208, 2022. doi:10.1109/ICSC52841.2022.00041.

- [19] Y. Tang, W. Wang, "A literature review of personalized learning algorithm", *Open Journal of Social Sciences*, vol. 6, pp. 119–127, 2018. doi:10.4236/jss.2018.61009.
- [20] Y.T. Zhuang, F. Wu, C. Chen, Y.h. Pan, "Challenges, and opportunities: from big data to knowledge in AI 2.0", *Frontiers of Information Technology & Electronic Engineering*, vol. 18, pp. 3–14, 2017. doi:10.1631/FITEE.1601883.
- [21] Z. Qiu, X. Wu, W. Fan, "Automatic distractor generation for multiple choice questions in standard tests", *28th Int. Conf. on Computational Linguistics, Int. Committee on Computational Linguistics, Barcelona, Spain*, pp. 2096–2106, 2020. doi: 10.18653/v1/2020.coling-main.189.
- [22] Z. Swiecki, H. Khosravi, G. Chen, R. Martinez-Maldonado, J. M. Lodge, S. Milligan, N. Selwyn, D. Gašević, "Assessment in the age of artificial intelligence", *Computers and Education: Artificial Intelligence*, vol. 3, 100075, 2022. doi: 10.1016/j.caeai.2022.100075.