# Fostering Critical Thinking in On-Demand Learning: Question-Based Support System with Slide Aids

**Saki INOUE** *

**Graduate School of Policy Studies**
**Kwansei Gakuin University**

**Abstract:**

In recent years, online learning has gained popularity in Japan, increasing opportunities for remote study using lecture slides. However, a common issue in Japanese education is that students tend not to ask questions in class. To address this issue, we propose an online learning support system that facilitates the process of students' questioning and thinking, referring to the research on "critical thinking". This study aims at clarifying the relationship between the questions generated by students and the class slides. This is achieved by analyzing the hierarchical relationship of the keywords that appear in the slides. We also propose a method of recommending appropriate lecture slides for the questions asked, which in turn, will encourage students to ask meaningful questions and engage in productive discussions.

**Key words and phrases:** e-learning, question generation, lecture slide recommendation, critical thinking, online learning support

## 1. Introduction

Introspective thinking, such as critical thinking, involves intentionally evaluating one's own thought processes. King claimed that good thinkers ask good questions (King, 1995). However, it is common for Japanese students to be too hesitant to query the lecture content. For instance, Ikuta et al. observed that 47.7% of youngsters did not ask or generate inquiries (Ikuta and

* Graduate School of Policy Studies Kwansei Gakuin University（hms54907@kwansei.ac.jp）

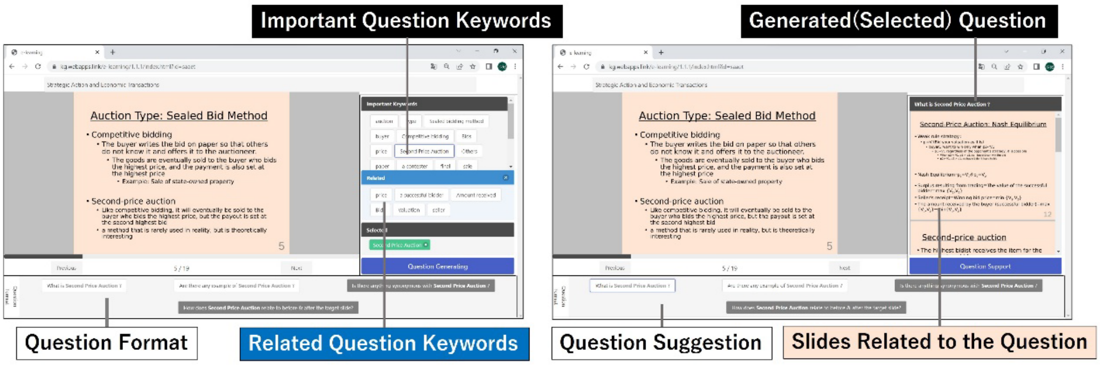


Figure 1 Support for generating questions (left) and presenting relevant slides (right) for critical thinking.

Maruno, 2004). Therefore, the purpose of this study was to encourage students to create questions and to support the process of careful consideration. Figure 1 illustrates the UI of the implemented learning support system. Depending on the behavior of the learner, our proposed UI switches from “question generation support” to “presentation of relevant slides.” First, we provide a question suggesting support task, where the learner's current “target slide” is shown in the upper-left corner of the UI. According to the proposed approach, the target slide's keywords are also shown in the upper-right corner. Additionally, when a learner selects the “Important Keyword Suggestion” option, the keywords on the target slide are highlighted in accordance with their importance so that they are easier comprehended.

## 2. Overview

Figure 3 illustrates the schematic overview of the system in this study. Instructors share their crafted lecture slides with students and store them in a database named Lecture Slide DB. As users of this system, learners utilize it to watch class videos presented by instructors. While viewing these videos, learners actively generate questions concurrently with the displayed question keywords and formats. Subsequently, the system conducts exploration to find slides corresponding to the questions and offers slide recommendations in a ranked format. The proposed ranking method utilizes the question keywords and formats chosen by learners, in addition to the hierarchical structure of slides created by instructors.

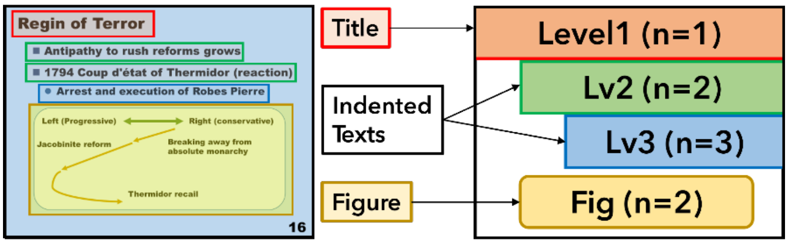


Figure 2 Examples of indentation and hierarchy levels

## 3. Support for Question Generation

PowerPoint and Keynote slides used in lectures typically feature a hierarchical structure with indentation.

### 3.1 Important Question Keyword Extraction

We suggest utilizing the slides' hierarchical structure to prioritize query keywords according to importance. The indentation is assigned the hierarchical levels we defined, which reprocesses the slide's hierarchical structure. The title has the highest hierarchy level ($n=1$), as shown in Figure 2, and after that, the hierarchy level rises one level in descending order. First, we perform a morphological analysis on the target slide and extract nouns that might be used as question keywords. The importance of the keyword $k$ in Slide $x$ is calculated by considering the frequency of the keyword's occurrence on the slide and the height of the occurrence's hierarchical level. In this paper, I hypothesize that cases where the keyword appears frequently on the slide and occurs at a higher hierarchical level are more important as question keywords.

### 3.2 Keyword Extraction for Related Questions

Next, while choosing keywords, learners could decide to choose relevant terms that do not exist on a target slide. By computing the relationship between the keywords in the material and figuring out the upper-lower relationship between the chosen and other keywords, we respond to this request in this case. In this work, we use the hierarchical structure of the lecture slides to extract a thesaurus from them. This enables learners to choose relevant keywords that appear outside of the target slide.

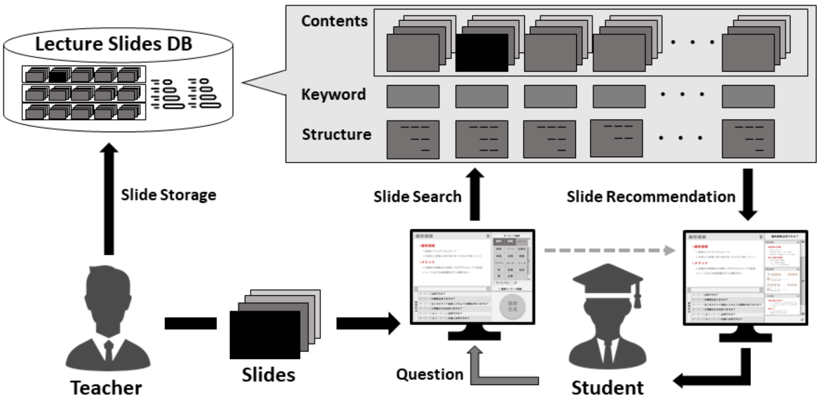


Figure 3 System overview

### 3.3 Presentation of Question Format

We provide a format for questions that are created by simply combining keywords. The question stem list produced by Ikuta et al. (Ikuta and Maruno, 2004) and the question categorization approach given by Shinogaya (Shinogaya, 2013) were used as the basis for the question format presented in Table 1. In this work, we define high-level questions as those that promote linkages between keywords and prior knowledge, whereas low-level questions are those that confirm the facts of the keywords themselves. Additionally, we consider the situation when learners choose one term for a single keyword question and various keywords for a multiple-keyword question. Both $Q_{L_3}$ and $Q_{H_3}$ allow various keywords in the question format.

The slide rank computation for the question in this study uses the question format. The goal of this study is to support learners in generating and reflecting on possible high-level questions. This is because low-level questions can ultimately be solved with simple methods such as web search or by browsing Wikipedia. We selected three question patterns (six in total) for evaluation, which are assumed to be low-level and high-level.

Table 1 Question format

| | |
|---|---|
| Low | $Q_{L_1}$: What is $k_1$? |
| | $Q_{L_2}$: Are there any examples of $k_1$? |
| | $Q_{L_3}$: What are $k_1$ *or* $k_2$? |
| High | $Q_{H_1}$: Is anything synonymous with $k_1$? |
| | $Q_{H_2}$: How does $k_1$ relevant to before and after the target slide? |
| | $Q_{H_3}$: What is the difference between $k_1$ *and* $k_2$? |

## 4. Slide Recommendations for based on Question Format

In our study, we categorized questions into low- and high-level types to align them with suitable slide categories, subsequently displaying relevant slides to users. We defined four slide types based on keyword appearance positions and hierarchy relationships to characterize recommended candidate slides for specific questions. Notably, question formats with multiple keywords were not addressed in this paper, leaving room for future exploration.

### 4.1 Determination of slide types

To characterize the recommended candidate slides for a specific question, we defined four slide types based on the appearance position and hierarchy relationships of the keywords in the question. Based on the indent level of target slide, there are two types of slides: higher and lower slides. Higher slides provide an earlier direct explanation of the question keyword, while lower slides supplement the explanation by addressing the question keyword later. Next, we extract the relationship between the keyword semantic hierarchy and the slide hierarchy. This represents a partial order relationship between question keywords and other keywords.

Knowledge Hierarchy : onion ≺ vegetable, tomato ≺ vegetable, vegetable ≺ food, vegetable ≺ green

| Slide Type | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Example $Q_k$=“*vegetable*” | vegetable<br>onion<br>tomato<br>tomato | vegetable<br>food<br>green<br>green | food<br>green<br>food<br>vegetable | onion<br>onion<br>tomato<br>vegetable |
| Indent Level | higher | higher | lower | lower |
| Knowledge Relationship | normal | reverse | normal | reverse |
| Question Format | $Q_{L_1}, Q_{L_2}, Q_{L_3}$ | $Q_{H_1}, Q_{H_2}, Q_{H_3}$ | $Q_{H_1}, Q_{H_2}, Q_{H_3}$ | $Q_{L_1}, Q_{L_2}, Q_{L_3}$ |

figure 4 Slide type categorization based on description tendencies and keyword hierarchy.

For example, if a keyword is a more specific instance of a question keyword, it is expressed as a hyponym of the question keyword. We examined 170 pages of lecture slides from four courses and identified specialized or generalized roles for each slide type. Types (1) and (4) were determined to play specialized roles, while types (2) and (3) played generalized roles. Figure 4 provides examples of the four slide types.

## 4.2 Grouping of Recommended Slides

The grouping of slides recommends slides that have no occurrence of question keywords adjacent to the recommendation candidate slides. We considered three types of slide grouping methods:

A. Grouping of adjacent similar slides

B. Grouping candidate slides by glue bonding

C. Group similar adjacent slides (considering hierarchical)

| Lecture ID | Slide Page | Proposed Method | | | Topic |
|---|---|---|---|---|---|
| | | Proposal 1 ($\omega_1 = 0.5$) | Proposal 2 ($\omega_1 = 0.8$) | Proposal 3 ($\omega_1 = 0.2$) | |
| 1 | 11 | 0.845 | 0.868 | 0.807 | **0.919** |
| | 16 | 0.625 | 0.625 | 0.648 | **0.870** |
| | 19 | 0.877 | 0.794 | 0.883 | **0.995** |
| 2 | 8 | 0.975 | 0.975 | 0.975 | **0.996** |
| | 10 | 0.905 | 0.822 | 0.905 | **0.959** |
| | 17 | 0.695 | 0.695 | 0.693 | **0.839** |
| 3 | 25 | **0.875** | **0.875** | 0.854 | 0.840 |
| | 26 | 0.647 | 0.647 | 0.647 | **0.883** |
| | 31 | 0.722 | 0.709 | 0.845 | **0.850** |
| 4 | 3 | 0.980 | 0.943 | 0.983 | **0.995** |
| | 11 | **0.929** | **0.929** | **0.929** | **0.929** |
| | 14 | 0.646 | 0.646 | 0.646 | **0.995** |
| All | Ave | 0.810 | 0.794 | 0.818 | 0.923 |
| | Rank | 2 | 3 | 1 | — |

| Lecture ID | Slide Page | Conventional Method | | | Proposed |
|---|---|---|---|---|---|
| | | TF | Total Hierarchy | Average Hierarchy | Proposal ($\omega_1 = 0.5$) |
| 1 | 11 | 0.619 | 0.906 | **0.910** | 0.906 |
| | 16 | 0.761 | **0.800** | 0.740 | 0.777 |
| | 19 | 0.822 | **0.898** | 0.799 | 0.891 |
| 2 | 8 | 0.804 | **0.991** | 0.986 | **0.991** |
| | 10 | **0.946** | **0.946** | 0.857 | **0.946** |
| | 17 | 0.768 | 0.805 | 0.726 | **0.809** |
| 3 | 25 | 0.759 | 0.790 | 0.790 | **0.828** |
| | 26 | 0.525 | **0.581** | 0.581 | **0.581** |
| | 31 | 0.593 | **0.798** | 0.715 | 0.725 |
| 4 | 3 | 0.972 | **0.988** | 0.909 | 0.985 |
| | 11 | **0.958** | **0.958** | 0.931 | **0.958** |
| | 14 | 0.506 | **0.698** | **0.698** | **0.698** |
| All | Ave | 0.753 | **0.846** | 0.803 | 0.841 |
| | Rank | 4 | 1 | 3 | 2 |

Figure 5 Tuning effects of the proposed methods (left) and Comparison of conventional methods and Proposal 1 (right)

## 5. Evaluation

We provided three preliminary questions to assess the participants' comprehension levels of the lecture subjects. First, we asked about their prior learning experience related to the subject matter addressed in the lectures. Second, we inquired about their self-reported understanding of the lecture subjects. Finally, we conducted a mini-test specifically focused on the subjects of the lectures. These questions were designed to assess the participants' comprehension levels regarding the content presented in the experimental lecture slides. As a result, it was found that there were variations in students' level of learning experience and confidence in their understanding depending on the content of the lectures, and it was found that this point should be kept in mind in future evaluations.

### 5.1 Experiment 1

In Experiment 1, we evaluated the rankings using the normalized discounted cumulative gain (nDCG) metric. The correct answer data is derived from the ranking of keywords that participants want to ask questions about, as indicated in their responses. This allows us to assess the effectiveness of the ranking method in accurately identifying keywords of interest for each participant. We present the results of how appropriate the topic keywords answered by the participants are as question keywords in Figure 5. The ``Topics" column represents the rank correlation between the ranking of the correct answer data (i.e., the ranking of the keyword to be asked) and the topic ranking of the corresponding keyword as answered by the participant. The average value is *0.923*, with a minimum of *0.839* and a maximum of *0.996*, indicating a high level of agreement.

## 5.2 Experiment 2

In our evaluation, we compared and assessed the performance of three conventional methods and the proposed method using the rank correlation of nDCG. The following three methods were evaluated as conventional methods:

- Number of occurrences of the keyword
- Total Hierarchy: total reciprocal of keyword occurrence hierarchy level
- Average Hierarchy: Value obtained by dividing Total Hierarchy by TF

We present the evaluation results comparing topic keywords' suitability between the proposed and conventional methods in Figure 5. The Total Hierarchy method achieved the highest accuracy, with the average value of *0.846*. Following in second place was Proposal 1 (baseline) with an average value of *0.841*. The Average Hierarchy method ranks third with an average value of *0.803*. The TF method exhibits the lowest average value among the tested approaches, measuring *0.753*. We evaluated how tuning the calculation of question keywords using the proposed method affects the appropriateness of the question keywords. The results are presented in Figure 5. We conducted a validation to assess the effectiveness of adjusting the weights of keyword slide importance, as described in Section 3.2.1. The baseline is Proposal 1, which represents the proposed method without any weight adjustments and has $w_1$=*0.5*. Proposal 3 achieved the highest accuracy, with an average value of *0.818*. Proposition 1 (Baseline) was second, with an average value of *0.810*. Proposal 2 had the lowest mean value among the tested approaches, at *0.794*.

## 5.3 Experiment 3

We conducted a preliminary experiment and survey to verify the slide types. The subjects were 9 university students in their twenties. We conducted a preliminary experiment using 170 slides from four lectures their theme is vegetable to investigate the frequency of appearance of each slide type. Types (1) and (3) exhibited higher frequencies than types (2) and (4). Lecture C, which focused on ``tomatoes'', showed a different trend, with many reverse slides. These findings align with the authors' expectations.

To verify the effectiveness of matching slide types and question formats, we examined the percentage of slide types judged to be appropriate based on the top-*50%* ranking of the proposed method. For $Q_{L_1}$, slide types 1 and 4 together accounted for *56%* of the top ranking, indicating little influence. For $Q_{L_2}$, slide types 1 and 4 accounted for *60%* of the data, and for $Q_{H_1}$, slide types 2 and 3 accounted for *67%* of the data, showing slightly higher ratings. The results revealed that matching the question format with the four slide types had little impact because of the insufficient matching between the slide type classification, which focused on the partially ordered relation between keywords and the question format.

## 6. Conclusion

In this study, we have proposed an online learning support system to promote critical thinking in students and performed an experiment to evaluate the proposed approach. This study focused on developing an online learning support system to enhance critical thinking skills in students. Overall, our study provides valuable insights for the development of effective online learning support systems. In future research, we shall conduct large-scale evaluation experiments to investigate the extent to which the proposed method can stimulate critical thinking. We plan to expand the method of determining slide types to enable the system to handle questions that contain multiple keywords. Ultimately, our objective is to implement the system in real-world classrooms and conduct both quantitative and qualitative evaluations of the number and quality of generated questions. In the future, we may expand the target lecture content to include media such as lecture videos, audio lessons recorded by teachers, and other educational content.

## References *Cited references in this paper. Other references are listed in master thesis.

＜英語＞（ABC 順）

King. Designing the instructional process to enhance critical thinking across the curriculum: Inquiring minds really do want to know: Using questioning to teach critical thinking. Teaching of Psychology, 22(1):13–17, 1995.

J. Ikuta and S. Maruno. Does the child generate a question in elementary school class? : Relation between interrogative feeling and the question generation, and expression. Psychological Research of Kyushu University, 5:9–18, 2004.

K. Shinogaya. An examination of the impact of interventions in question generation and answer generation during preparatory activities on class comprehension and its process [in japanese]. Educational Psychology Research, 61(4):351–361, 2013.

# オンデマンド授業における学生の批判的思考を促進する学習支援システム

井上　沙紀 *

【要旨】

遠隔・オンライン教育の急速な普及により，ポストコロナ期にも対面授業だけでなく，遠隔・オンライン教育の利点を活用した教育が求められている．しかし，教師との意思疎通の難しさや一方向の講義スタイルであることが懸念されている．特に，授業動画を配信するオンデマンド型授業では，教員が学生の理解度や状況を把握できないために学習支援が難しいという課題がある．そこで我々は，よき思考者はよき質問者であるという批判的思考の立場から，学生の質問応答を支援するシステムの開発を提案してきた．本稿では，前年度までの学生の質問データと講義スライドの比較を行うことで，学生に対して質問キーワードや質問形式の候補を推薦する手法を提案する．

キーワード：e-Learning，質疑応答，相互評価，学習支援

* 関西学院大学大学院総合政策研究科博士課程前期課程（hms54907@kwansei.ac.jp）