Integrated Science Education Journal
Vol. 5, No. 2, May 2024, pp. 91~100
ISSN: 2716-3725, DOI: [10.37251/isej.v5i2.1003](https://doi.org/10.37251/isej.v5i2.1003)

Development of Logic-Based Question Classification Methods to Measure the Ability to Ask Questions Essential for Setting Research Questions

Li Qiu[1], Fumihito Ikeda[2], Naoko Yamashita[3]
[1,3]Department of Natural History Science, Graduate School of Science, Hokkaido University, Hokkaido, Japan
[2]Faculty of Liberal Arts, Science and Global Education, Osaka Metropolitan University, Osaka, Japan

Article Info

Article history:

Received Apr 03, 2024
Revised Apr 22, 2024
Accepted May 20, 2024
OnlineFirst May 31, 2024

Keywords:

Logical Reasoning
Question Perspectives
Research Questions
Scientific Inquiry

ABSTRACT

Purpose of the study: Many learners experience difficulties in generating questions that lead to research questions. This study aimed to present perspectives on questions that inquire into the logic within each type of logical reasoning, propose a new question classification method based on these perspectives, verify its objectivity, and examine participants' logicality.

Methodology: This study deductively developed perspectives and classification methods for questions based on the logical reasoning theory. To verify the question-classification method and examine participants' logicality, 1,164 Self-described questions from 24 graduate-level students were collected. These questions were classified by two raters using qualitative research software, followed by a statistical analysis. Descriptive and inferential statistics were used for the analysis.

Main Findings: Based on the analysis of the question data, the degree of agreement on question classification between the two raters was high. Notably, there was exceptional concordance in categorizing questions asking about deductive reasoning, hypothetical reasoning, and specific questions. Additionally, the analysis of question characteristics revealed a predominant presence of questions asking about inductive reasoning and ambiguous questions. In contrast, there were fewer questions asking about deductive reasoning, hypothetical reasoning, and specific questions.

Novelty/Originality of this study: This research adds a logical perspective to traditional question frameworks and develops objective question classification methods. It provides a framework to support the formation of questions that lead to research questions and a method to objectively assess the quality of learning question formation techniques, with significant implications for educational practices. This study is limited by few raters, few specific questions.

Corresponding Author:
Fumihito Ikeda,
Faculty of Liberal Arts, Science and Global Education, Osaka Metropolitan University,
Osaka 558-8585, Japan
Email: fumike@omu.ac.jp

1. INTRODUCTION

We are currently facing various challenges such as climate cries, the proliferation of misinformation, and the rapid expansion of digital technologies [1]. These issues have ushered us into a VUCA era, which characterized by volatility, uncertainty, complexity, and ambiguity [2]. In such times, learners are increasingly expected to utilize science proactively to advance science and technology, drive societal change, and create new value [1].

Questions are important tools for stimulating creativity and generating new ideas and knowledge [3]. Therefore, questions are the starting point of scientific discovery and play a key role in the processes of inquiry and research [4], [5]. However, many questions generated by learners tend to remain at the level of confirming facts and procedures [6]. In other words, learners find it difficult to spontaneously generate questions that can evolve into research questions and potentially advance science and technology. Therefore, unlike authentic scientific inquiry where scientists generate their own questions and explore the answers, in most school-based inquiry activities, the research questions are directly provided to the learners [7].

Wittgenstein [8] revealed that the totality of facts forms the world, asserting that "logic fills the world: the limits of the world are also its limits." This can be interpreted as the limitations of science, which is a part of the world determined by logic. Popper argued that scientific theories should be falsifiable and presented this as a criterion to distinguish between science and non-science, characterizing falsifiability through "the logical relationships that exist between the theory and the class of basic propositions." He revealed that the development of science is a process of constantly testing existing theories and adding new theories when faced with facts that cannot be handled, thereby improving or updating the reliability of theories [9]. According to Popper, scientific development is achieved by critically questioning the logic between theories and the groups of propositions that establish those theories. Kuhn pointed out that science does not always develop continuously based on existing paradigms and that "anomalies" that cannot be explained by existing paradigms often exist, arguing that the development of science progresses through paradigm shifts [10]. In the process of continuously testing the existing paradigm, when difficult-to-explain cases emerge, the old paradigm may gradually be reconsidered, and a new paradigm that can adequately explain these anomalies may emerge. Therefore, the requestioning of the logic between the existing paradigm and its propositions becomes a foundation for paradigm shifts. In the scientific field, questioning the logic between the existing theories and the group of propositions that compose these theories is an important element for breaking the limits of science and promoting the development and innovation of science and technology. Consequently, It is important for learners generate questions that probe logic forming the basis of research questions, in order to engage in authentic scientific inquiry and create new value.

Providing a framework for questions facilitates the generation of relevant questions within its scope indicated by the framework [11]. Moreover, to enhance the quality of education and measure the achievement of learning objectives in science education, it is essential to objectively assess the learning processes and outcomes of learners in the classroom[12], [13]. Therefore, to effectively form research questions , it is important to identify the types of questions that probe logic and clearly present their schema ,and it is crucial for learners to accurately understand question formation methods and the quality of their learning, which necessitates an objective classification method for questions.

But, Questions that lead to research questions have been primarily classified based on their form and content. From a formal perspective, De Vaus categorized questions that can be posed in social science research into two types based on the use of interrogatives: a "what" type concerning the description of reality, and a "why" type pertaining to the explanation of mechanisms [14]. To further develop De Vaus's classification, White used the commonly employed journalistic 5W1H (Who, What, When, Where, Why, How) to further subdivide questions related to description and explanation, stating the research methods corresponding to each question type [15]. From the content perspective, Aristotle typified questions that could be raised in inquiry as those concerning the existence, nature, attributes, and causes of an object [16]. Inspired by Aristotle's classification, Dillon divided questions that can be posed in scientific inquiry and research based on the types of propositions (knowledge) obtained about a phenomenon into "primary questions" asking about basic attributes, such as the existence and function of a phenomenon; "secondary questions" regarding comparative attributes, such as relationships, similarities, and differences between two or more phenomena; "tertiary questions" asking about the relationship between two phenomena or objects; and "other questions" asking about new methods of understanding the phenomenon [17]. Moreover, Chin and Kayalvizhi proposed a classification of investigable questions regarding the characteristics of a phenomenon or the relationship between variables and non-investigable questions related to religion, among others [18]. These classifications help to grasp the scope of questions in scientific inquiry and research, recognize the direction of questions about the subject, and identify missing content, thereby aiding in the generation of questions that can evolve into research questions.

From a logic perspective, [19] proposed the existence of "Questions to Close" in scientific inquiries, which questions the veracity of conditional propositions or the unexamined conditional propositions that extend scientific knowledge. This is because scientific knowledge consists of propositions that are inductively or deductively falsifiable, and the falsifiability of propositions and the conditional propositions that constitute them can be considered. However they pointed out that this alone could not facilitate the application of scientific knowledge or promote paradigm shifts. Thus, "Questions to Extend," which inquire about propositions that can be deductively derived by assuming certain propositions to be true, and "Questions to open," which question the propositions, are necessary [19]. In this study, within the category "Question to Close," the focus is on the logic between the theory and the conditional propositions that establish it, presenting a framework for questions related to conditional propositions. Like this, these studies have focused on the types of propositions obtained by answering questions overlooking the aspect of logic. Among them, [19] classified scientific questions from a logical perspective, but a detailed examination regarding specific question perspectives that probe the logic between existing theories and their supporting propositions in each logical reasoning as well as question classification methods based on these perspectives is lacking.

Therefore, this study formed the following research questions: What categories of questions that probe logic exist, and how can questions be objectively classified based on logic? What will be the logicality of the participants? This study first proposes perspectives on questions that probe the logic within each logical reasoning and introduces a new classification method for questions based on these perspectives while simultaneously verifying its objectivity. Second, this study seeks to reveal the logicality of participants through the questions they posed. This provides a framework for generating questions that lead to research questions , which are essential for learners to conduct authentic scientific inquiry. It is hoped that this study will objectively grasp the logicality of learners, provide a foundation for educational interventions, and understand the learning situation of question generation techniques, giving direction for learning improvement.

2. RESEARCH METHOD

Following Barton's approach [20], we established classification axes based on theories of logical reasoning and deductively constructed perspectives and classification methods for questions. Logical reasoning can be classified into three forms: inductive, deductive, and hypothetical. This division is based on the inference of one of the elements, that is, premise; rule; or result, from the other two [21]. Additionally, as demonstrated in Michita [22], the completeness of inference elements is the first condition for logic, and the validity of the relationships between elements is the second condition. The questions concerning logical reasoning are based on these conditions, with the first-level questions directly probing inductive, deductive, and hypothetical reasoning. Second-level questions were established to break down the first-level questions and explore the relationships between elements. The second-level questions, namely, questions for logical reasoning, can be considered from three perspectives within the reasoning process. Perspective 1 probes the set of propositions that form the starting point of the inference, Perspective 2 questions the validity of the connections between each proposition, and Perspective 3 queries the process of inferring a conclusion from these sets of propositions (Figure 1).

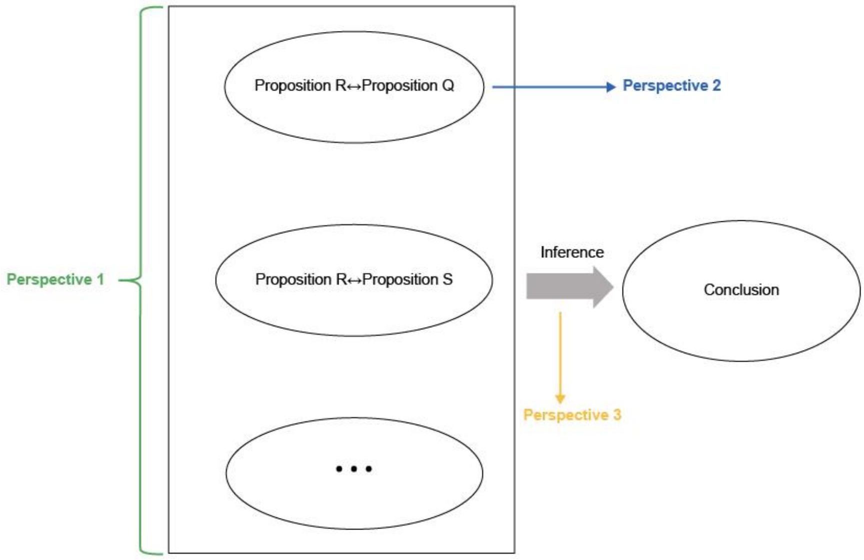

Figure 1. Perspectives on questions for logical reasoning

The questions regarding logical reasoning were as follows. First, the questions regarding inductive reasoning are explained. Inductive reasoning is a method that abstracts from multiple premises and results and derives general rules. The first question is whether a specific general law can be derived from a specific set of premises and results. The second level of questions, which breaks down this first-level question, includes the following three perspectives. The first involves questions about the set of premises and results that form the basis of inductive reasoning, specifically, whether a limited set of premises and results is sufficient to derive a rule. The second perspective verifies whether the result can always be obtained based on these premises, that is, the possibility of deficiencies in the premises that affect the result. The third perspective determines the validity of information omissions and the presence of distortion when deriving a conclusion from the set of premises and results.

Deductive reasoning involves applying a rule to a certain premise and predicting the results. The first-level questions ask whether a result can be predicted based on a given premise and rule. Deductive reasoning has the characteristic that if the premises and rules are true, the conclusion is necessarily true. However, scientific inquiry and research often involves implicit premises and rules, leaving room to doubt the validity of the logic. The second-level questions can be considered from three perspectives, similar to inductive reasoning. The first perspective investigates whether a set of premises and rules is sufficient to lead to a result. The second perspective questions the relationship between the premise and the rule, that is, whether the rule applied to the premise is truly valid. The third perspective examines the validity of information omission and the presence of distortion in the process of predicting results based on a set of premises and rules.

Hypothetical reasoning is a method that applies a rule to a specific result to infer the premises that led to that result. First, a first-level question applies a rule to an observed result and asks about the possible premises. When inferring the premises that lead to a result, a possibility that numerous rules could be applied to that result exists; therefore, similar to inductive and deductive reasoning, second-level questions also have three perspectives. These perspectives question whether the set of results and rules is sufficient to infer the premises, the validity of the rules applied to the observed results, validity of information omission, and presence of distortion in the process of inferring. The three perspectives on the questions for logical reasoning are presented in Table 1.

Table 1. Perspectives on questions for the three types of logical reasoning

A	Perspectives on questions for inductive reasoning	
	1.	Questioning the validity of the set of premises and results
	2.	Questioning the validity of the relationship between individual premises and results
	3.	Questioning the validity of information omission and the presence of distortion during the process of generalization
B	Perspectives on questions for deductive reasoning	
	1.	Questioning the validity of the set of premises and rules
	2.	Questioning the validity of the relationship between individual premises and rules
	3.	Questioning the validity of information omission and the presence of distortion during the process of prediction
C	Perspectives on questions for hypothetical reasoning	
	1.	Questioning the validity of the set of results and rules
	2.	Questioning the validity of the relationship between the results and individual rules
	3.	Questioning the validity of information omission and the presence of distortion during the process of inference

The question classification method based on the perspectives of logic described above was developed around the three common elements of premises, rules, and results, which are integral to inductive, deductive, and hypothetical reasoning. The specific steps are as follows.

(1) Element identification: Elements corresponding to the premises, rules, and results in a question statement are identified.

(2) Classification of questions: Based on the identified elements, it is determined which of the following five types each question belongs to, and the appropriate type is assigned.

a) Ambiguous questions
Those that contain only one of either the premises or results.

b) Questions asking about inductive reasoning
Those that contain premises and results and inquire about a rule that can be generalized.

c) Questions asking about deductive reasoning
Those that contain premises and rules and inquire about a predictable result.

d) Questions asking about hypothetical reasoning
Those that contain results and rules and. inquire about an inferable premise.

e) Specific questions
Those that inquire from the perspectives shown in Table 1, namely those that include all elements of premises, rules, and results and further inquire about one of these elements.

Types **b), c),** and **d)** correspond to first-level questions that directly probe logical reasoning, whereas type **e)** corresponds to second-level questions that examine the validity of logical reasoning.

This research employs a quantitative approach, which is considered appropriate for verifying the objectivity of the proposed question classification method and for examining the current state of participants' logicality. The quantitative method, suited for testing theory and hypothesis, is considered more " objective" due to its quantitative nature and is frequently used to understand the current state [23].

When researchers have limited resources, they can use the convenience sampling technique, which allows them to select samples that are easily accessible [24]. Thus, considering the feasibility of the study , this approach was adopted to select the sample. The study participants 24 students who were enrolled in the course "Socio-constructivism for Science and Technology I," a specialized class in the Faculty of Science and an Inter-graduate School Class for graduate students at H University in Japan, during the spring term of 2020 (April to May). All participants were first-year master's students. The reasons for selecting graduate students were as follows. Many university graduates and new graduate students are reported to have improved their research skills, including the discovery and formulation of problems, due to their research experience [25].

The course consisted of seven lectures, each lasting 90 minutes, presented along the following seven themes: the importance of questions and factors that influence curiosity, human perceptual characteristics, human visual characteristics, characteristics of human consciousness and attention, the importance of better questions, the definition of science, and science from the perspective of social constructionism.

The mothed of self-describing questions on a sheet is an effective technique for collecting questions [26]. Therefore, this study adopted this method, and the survey sheets were designed to consist of two stages, based on the theory that breaking down the questions can further refine the content [27]. This approach is believed to enhance the likehood of recognizing the validity of the relationships between the elements contained in the question text. After each lecture, participants filled out these survey sheets. They were encouraged to pose numerous questions. In the first stage, the participants were prompted to raise questions related to the lecture content (referred to as the initial question). Next, from among these initial questions, they were asked to select the one they considered the most important, explain why, and then raise further questions to delve deeper into their chosen question (referred to as follow-up questions). Through the email client and learning management system, which was accessible to the faculty and students, questions related to each lecture's content were submitted in the following class. In the first class, it was explained that the data would be used solely for research purposes, and consent from the participants was obtained on this basis.

Objectivity refers to the fair property that ensures that the results match when different people classify and evaluate the same subject [28]. Moreover, background knowledge about the subject and method is important for the accuracy of the classification work [29]. Therefore, to validate the objectivity of the proposed classification method, a graduate student from the same institution as the first author classified the questions independently. In the analysis of qualitative data, coding is primarily done using qualitative data management software [30]. This classification was conducted using the qualitative research software MAXQDA 2022, which is a convenient tool suitable for labeling and classifying data, allowing for the analysis through a series of processes including data organization and coding settings [31]. Drawing from [30], [31], the specific procedures were as follows: **(1) Importing and organizing the question data:** All questions were imported and checked for any that were not in the form of questions, and inappropriate questions were excluded.; **(2) Setting question classification codes:** To identify ambiguous and first-level questions, main codes, such as "elements of inference," "ambiguous questions," "asking about inductive reasoning," "asking about deductive reasoning," and "asking about hypothetical reasoning," were set. In addition, to identify specific inference elements and second-level specific questions, sub-codes, such as "premise," "rule," "result," "set of premises and results," "relationship between premise and result," and "generalization causing information omission/distortion," were assigned to "inductive reasoning." Similar sub-codes were also assigned to "deductive reasoning" and "hypothetical reasoning" to provide the necessary codes for question classification.; **(3) Sharing files and explaining the operational methods:** Project files containing questions and question classification codes were shared among the raters. Furthermore, the first author explained in detail the question classification method and application of codes using each type of question example generated based on the lecture content, which was different from those asked by the participants, and conducted a classification exercise. For example, the question "What is the relationship between language and ways of illusion?" identified "language and ways of illusion" as premise, to which a "premise" code was assigned. "Has a relationship" was identified as the result, and it was thus assigned the code for "result." Considering that it enquires about the rule deriving from the language and ways of illusion, the "asking about inductive reasoning" code was assigned. Based on the above operational methods, each question was individually classified, and the appropriate codes were selected.

The agreement between the raters and the number of questions of each type obtained through these classification activities were analyzed using SPSS 29. This analysis included descriptive statistics summarizing the number of questions by type as well as inferential statistics to test inter-rater agreement and differences in the number of questions across types. To evaluate the inter-rater agreement, Cohen's kappa coefficients were utilized. This coefficient is widely used across various fields to measure the degree of agreement between different raters when classifying the same object [32]. To compare the differences in number of questions across types, a chi-square goodness-of-fit test was conducted. This allowed for the determination of statistically significant differences among the various types of questions. Through this analysis, the study aimed to gain insight into the objectivity of the question classification method and participants' logicality.

3. RESULTS AND DISCUSSION

3.1. Inter-rater agreement on question classification

Invalid questions were excluded, resulting in 1.164 questions raised by participants during the seven lectures having been analyzed. Specific examples of questions agreed upon by the two raters and their classification results are shown in Table 2.

Table 2. Examples of consensus in question classification among raters

Question	Classification result
How to ask high-quality questions?	Ambiguous question
What actions does one take when possessing intellectual curiosity?	Question asking about inductive reasoning
When illusions occur, people try to perceive Gestalt based on their knowledge and experience. Does the way they experience the illusion differ depending on their individual knowledge and experiences?	Question asking about deductive reasoning
The unknown is a source of anxiety and fear. Is our instinctive fear or aversion to things that are different from ourselves a result of ignorance?	Question asking about hypothetical reasoning
The fact that a hypothesis cannot be subjected to a falsification test does not mean that it can be definitively declared false. What other theories, such as string theory, cannot be falsified but cannot necessarily be considered untrue?	Specific question
Curiosity is influenced by the degree of surprise, the amount of knowledge, the level of confidence, and the complexity of the subject. However, can the extent of curiosity and its relationship with the amount of knowledge, the level of surprise, the strength of confidence, and the complexity of the subject be individually quantified?	Specific question

To verify the objectivity of the classification method for questions based on logical reasoning, the classification results of the questions among the raters were first cross-tabulated (Table 3). Then, based on the data from this cross-tabulation, Cohen's kappa coefficient was calculated for all questions. To analyze the degree of agreement for each type of question in detail, Cohen's kappa coefficients were calculated for each type. The results are summarized in Table 4. These results confirmed that the *k* coefficients for all the questions were high. In particular, the *k* coefficients for the questions on deductive reasoning, hypothetical reasoning, and specific questions were extremely high. The high *k* coefficients among the raters suggested that the question classification method based on the submitted perspectives was highly objective, eliminating personal biases or preferences of the raters and enabling consistent judgments. The level of agreement in the classification of questions asking about deductive reasoning, hypothetical reasoning, and specific questions was remarkably high, indicating that when applying the classification method to these questions, there was little difference in interpretation among raters, leading to more objective outcomes. However, since the occurrence of questions regarding deductive reasoning, hypothetical reasoning, and specific questions was relatively low in this classification task, caution should be exercised in generalizing the results.

However, the *k* coefficients for ambiguous questions and questions asking about inductive reasoning were relatively low. As shown in Table 3, it was common for one rater to classify a question as ambiguous whereas another classified it as requiring inductive reasoning. For example, consider the question, "What beneficial outcomes have arisen from the questions I posed?" In this case, one rater considered the question to be ambiguous, noting that only the premise "I posed a question" was clear. Another rater identified the premise as "I posed a question" and the result as "beneficial outcomes," interpreting this as a rule, namely a question requiring inductive reasoning. After a discussion between the two raters, the question was classified as ambiguous because it only assumed the premise of "I posed a question," and it was unclear whether it sought to know a rule deriving benefit from this premise or the beneficial outcomes of applying rules. It suggests that some ambiguous questions and questions asking about inductive reasoning could be easily confused. This confusion may stem from the structural similarities in these questions, causing differences in interpretation among raters. For instance, questions like "What beneficial outcomes have arisen from the questions I posed?" categorized as ambiguous, and "What actions does one take when possessing intellectual curiosity?" categorized as probing about inductive reasoning have different scopes of queried premises but similar structures, which may lead to potential confusion among raters.

Table 3. Cross-tabulation among the raters

			Rater B					
			Ambiguous question	Question asking about inductive reasoning	Question asking about deductive reasoning	Question asking about hypothetical reasoning	Specific question	Total
Rater A	Ambiguous question	Frequency	282	106	1	2	4	395
		Rater A	71.4%	26.8%	0.3%	0.5%	1.0%	100%
		Rater B	82.7%	15.6%	2.1%	4.8%	7.3%	33.9%
	Question asking about inductive reasoning	Frequency	57	564	8	3	7	639
		Rater A	8.9%	88.3%	1.3%	0.5%	1.1%	100%
		Rater B	16.7%	88.2%	16.7%	7.1%	12.7%	54.9%
	Question asking about deductive reasoning	Frequency	0	2	36	0	0	38
		Rater A	0.0%	5.3%	94.7%	0.0%	0.0%	100%
		Rater B	0.0%	0.3%	75.0%	0.0%	0.0%	3.3%
	Question asking about hypothetical reasoning	Frequency	2	5	3	37	0	47
		Rater A	4.3%	10.6%	6.4%	78.7%	0.0%	100%
		Rater B	0.6%	0.7%	6.3%	88.1%	0.0%	4.0%
	Specific question	Frequency	0	1	0	0	44	45
		Rater A	0.0%	2.2%	0.0%	0.0%	97.8%	100%
		Rater B	0.0%	0.1%	0.0%	0.0%	80.0%	3.9%
	Total	Frequency	341	678	48	42	55	1164
		Rater A	29.3%	58.2%	4.1%	3.6%	4.7%	100%
		Rater B	100%	100%	100%	100%	100%	100%

Table 4. Inter-rater agreement on classification

	Agreement rate	Kappa	Lower CI	Upper CI
Ambiguous question	62%	0.66	0.60	0.72
Question asking about inductive reasoning	75%	0.67	0.61	0.73
Question asking about deductive reasoning	72%	0.83	0.77	0.88
Question asking about hypothetical reasoning	71%	0.83	0.78	0.88
Specific question	79%	0.88	0.82	0.93
All questions	83%	0.70	0.66	0.74

3.2. Characteristics of the number of questions

3.2.1. Characteristics of all questions

To explore the logicality of the participants, all questions with discrepancies were discussed between two raters to reconcile their decisions. Subsequently, the total number of questions submitted, including the initial and follow-up questions, was analyzed in detail. First, a chi-square goodness-of-fit test was conducted on the total number of ambiguous questions, first-level questions probing inductive, deductive, and hypothetical reasoning and second-level specific questions collected over seven lectures. The results showed a significant difference in the number of questions among these types, $\chi 2$ (4) = 1320.70, $p < .01$. Additionally, a residual analysis revealed that the number of questions enquiring about inductive reasoning and ambiguous questions were higher than questions asking about deductive reasoning, hypothetical reasoning, and specific questions. The number of questions per type, relative frequencies, and adjusted residuals are listed in Table 5.

Table 5. Distribution by question type and results of residual analysis in all questions

	Frequency	Relative Frequency(%)	Adjusted residual
Ambiguous question	333	29%	6.57
Question asking about inductive reasoning	678	58%	29.18
Question asking about deductive reasoning	46	4%	-12.24
Question asking about hypothetical reasoning	50	4%	-11.98
Specific question	57	5%	-11.52
Total	1164	100%	

3.2.2. Characteristics of the initial questions

To examine the characteristics of the initial questions, a chi-square goodness-of-fit test was conducted on the numbers of ambiguous questions, questions asking about inductive reasoning, deductive reasoning, hypothetical reasoning, and specific questions among initial questions, resulting in $\chi 2$ (4) = 674.04, $p < .01$. This confirmed the presence of significant differences. Residual analysis indicated that, similar to the overall findings, there were more questions asking about inductive reasoning and ambiguous questions and fewer questions asking about deductive reasoning, hypothetical reasoning, and specific questions. The frequencies, relative frequencies, and adjusted residuals for each question type are presented in Table 6.

Table 6. Distribution by question type and results of residual analysis in initial questions

	Frequency	Relative Frequency(%)	Adjusted residual
Ambiguous question	173	26%	3.19
Question asking about inductive reasoning	386	57%	21.47
Question asking about deductive reasoning	40	6%	-8.22
Question asking about hypothetical reasoning	37	5%	-8.48
Specific question	43	6%	-7.96
Total	679	100%	

3.2.3. Characteristics of the follow-up questions

Next, a similar chi-square goodness-of-fit test was conducted on the number of ambiguous questions, questions regarding inductive reasoning, deductive reasoning, hypothetical reasoning, and specific questions among follow-up questions, yielding $\chi 2$ (4) = 662.06, $p < .01$, indicating a significant difference. Residual analysis showed that the number of questions asking about inductive reasoning was the highest, followed by ambiguous questions. Conversely, there were fewer questions on deductive reasoning, hypothetical reasoning, and specific questions. The frequencies, relative frequencies, and adjusted residuals for each question type are presented in Table 7.

Although it was hypothesized that the proportion of second-level specific questions would increase in the follow-up questions raised to further break down the selected base initial questions, the analysis of the actual questions revealed that ambiguous questions constituted approximately 30% of all questions, initial questions, and follow-up questions, whereas specific questions accounted for only 5%. This indicates that some participants in the course were not sufficiently aware of the relationship between the elements of logical reasoning (premise, rule, and result) and tended to ask illogical questions, implying a possible lack of the first condition of logicality.

Table 7. Distribution by question type and results of residual analysis in follow-up questions

	Frequency	Relative Frequency(%)	Adjusted residual
Ambiguous question	160	33%	6.40
Question asking about inductive reasoning	292	60%	19.80
Question asking about deductive reasoning	6	1%	-9.24
Question asking about hypothetical reasoning	13	3%	-8.53
Specific question	14	3%	-8.43
Total	485	100%	

Moreover, the prominent number of questions asking about inductive reasoning and fewer questions probing deductive or hypothetical reasoning suggest that many participants did not contemplate the relationship between premises and results but were keen to learn the rules (knowledge) and requested immediate answers.

Considering that Japanese learners are required to provide answers in regular tests [33] and that traditional classes are often teacher-led, directly instructing learners, it should be borne in mind that even at the graduate level, the mindset of requesting answers may strongly persist.

The implications of this research include proposed perspectives and objective classification methods for questions not only expand the knowledge about questioning but can also be applied in educational settings to efficiently analyze learners' question characteristics or logic, identify gaps in question generation, and aid in prompting questions that lead to research questions and conducting authentic scientific inquiry. Moreover, the discovery that even graduate students may lack logic underscores the need to build a solid foundation for improving logic and question formation. Teaching should not involve merely imparting existing knowledge but rather promote a deep understanding of the methods of deriving and applying knowledge or rules and provide practical experience in using all three types of logical reasoning: inductive, deductive, and hypothetical. Furthermore, in agreement with the view outlined in [34], it is necessary to establish an examination system that allows for the demonstration of the required skills. To achieve this, it is essential to construct a testing system that enables candidates to demonstrate their "ability to ask questions," which is vital for setting research questions.

This study has certain limitations. The low frequency of questions asking about deductive reasoning, hypothetical reasoning, and specific questions in the classification task as well as the fact that only two raters participated in the classification process, are limitations. To address these concerns, future studies should first refine the classification method by considering the characteristics of easily confused question types, thus enabling raters to better understand the intentions behind the questions. Additionally, educational programs that utilize the question framework provided by this study should be developed, incorporating perspectives and concrete examples of questions and further promoting the formation of questions that demand logic. Finally, it is necessary to increase the number of raters and revalidate the objectivity of the classification method.

4. CONCLUSION

This study proposes perspectives on questions that probe logic within logical reasoning and introduces a new method for classifying questions based on these perspectives. We attempted a detailed examination of the objectivity of this method and the logic of the participants based on these classifications. The results revealed the following insights. First, the high degree of agreement in question classification demonstrates the objectivity of the proposed method, suggesting that it effectively eliminates the subjectivity of the raters and can yield consistent outcomes across different raters. Furthermore, the results of the analysis of the various types of questions posed by the participants suggest that even graduate students were prone to asking illogical questions and tended to request answers (rules), indicating a lack of logicality.

ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI Grant Number JP20K20420.

REFERENCES

[1] L. Valladares, "Scientific literacy and social transformation," *Sci. Educ.*, vol. 30, pp. 557-587, Apr. 2021, doi: 10.1007/s11191-021-00205-2.
[2] N. Bennett and G. J. Lemoine, "What a difference a word makes: Understanding threats to performance in a VUCA world," *Business Horizons*, vol. 57, no. 3, pp. 311-317, Jun. 2014, doi: 10.1016/j.bushor.2014.01.001.
[3] C. Chin and J. Osborne, "Students' questions: A potentail resource for teaching and learning science," *Stud. Sci. Educ*., vol. 44, no. 1, pp. 1-39, Feb. 2008, doi: 10.1080/03057260701828101.
[4] R. D. Vale, "The value of asking questions," *Mol. Biol. Cell*, vol. 24, no. 6, pp. 680-682, Oct. 2017, doi: 10.1091/mbc.e12-09-0660.
[5] I. Yanai and M. Lercher, "What is the question?" *Genome. Biol.*, vol. 20, Dec. 2019, Art. no. 289, doi: 10.1186/s13059-019-1902-1.
[6] C. Chin and D. E. Brown, "Student-generated questions: A meaningful aspect of learning in science," Int. J. Sci. Educ., vol. 24, no. 5, pp. 521-549, May. 2002, doi: 10.1080/09500690110095249.
[7] C. A. Chin and B. A. Malhotra, "Epistemologically authentic inquiry in school: A theoretical framework for evaluating inquiry tasks," *Sci. Educ.*, vol. 86, pp. 175-218, Feb. 2002, doi: 10.1002/sce.10001.
[8] P. Wittgenstein and C. K. Ogden, *Tractatus Logico-Philosophicus*. New York, NY, USA: Dover Publications, 1998.
[9] K. Popper, *The Logic of Scientific Discovery*, 2nd ed. Evanston, IL, USA: Routledge, 2002.
[10] T. S. Kuhn, *The Structure of Scientific Revolutions*, 2nd ed. Chicago, IL, USA: University of Chicago Press, 1970.
[11] A. King, "Inquiry minds really to do want to know: Using questioning to teach critical thinking," *Teach. Psychol.*, vol. 22, no. 1, pp.13-17, Feb. 1995, doi: 10.1207/s15328023top2201_5.
[12] A. Astalini, D. Darmaji, Dwi Agus Kurniawan, H. Jaya, and Sri Muslimatul Husna, "Analysis of teacher responses to the use of web-based assessment to assess students' attitudes towards science subjects," *In. Sci. Ed. J*, vol. 3, no. 3, pp. 66-71, Sep. 2022, doi: 10.37251/isej.v3i3.282.

[13] H. Harizon, “Description of teacher responses to the implementation of student process skills portfolio assessment,” *In. Sci. Ed. J*, vol. 4, no. 3, pp. 128-134, Sep. 2023, doi: 10.37251/isej.v4i3.731.

[14] D. De Vaus, *Research Design in Social Research*. Newbury Park, CA, USA:Sage, 2001, P. 1-2.

[15] P. White, “What makes a research question,” in *Development Research Questions,* 2nd ed. London, UK: Palgrave Macmillam, 2017, PP. 56-59.

[16] Aristotle, *Aristotle’s Posterior Analytics.* 1901(Transl.: in E. S. Bouchier, Whitefish, MT, USA:Kessing Publishing, 2010)

[17] J. T. Dillon, “The classification of research questions,” *Rev. Educ. Res.*, vol. 54, no. 3, pp. 327-361, 1984, doi: 10.3102/00346543054003327.

[18] C. Chin and G. Kayalvizhi, “Posing problems for open investigations: What questions do pupils ask?” *Res. Sci. Technol. Educ.*, vol. 20, no. 2, pp. 269-287, Dec. 2002, doi: 10.1080/0263514022000030499.

[19] F. Ikeda, “What are scientific questions?” *Journal Higher Education and Lifelong Learning*, vol. 28, pp. 35-43, 2021,

[20] A. H. Barton, “The concept of property-space in social research,” In *The Language of Social Research*, P. F. Lazarsfeld and M. Rosenberg, Eds., New York, NY, USA: Free Press, 1955, pp. 40-53.

[21] C. S. Peirce, “Deduction, induction, and phypothesis,” *Popular Science Monthly*, vol. 12, pp. 470-482, 1878.

[22] Y. Michita, “What is logical thinking?” *Bulletin of College of Education., University of the Ryukyus*, vol. 63, pp. 181-193, 2003, doi: 20.500.12000/528.

[23] Y. P. Gnawali, “Use of mathematics in quantitative research,” *Ganeshman Darpan*, vol. 7, no. 1, pp. 10–15, Dec. 2022, doi: 10.3126/gd.v7i1.53528.

[24] E. Iiker, A. M. Sulaiman, and S. A. Rukayya, “Comparison of convenience sampling and purposive sampling,” *American Journal of Theoretical and Applied Statistics*, vol. 5, no. 1, pp. 1-4, Dec. 2015, doi: https://doi.org/10.11648/j.ajtas.20160501.11.

[25] J. Gilmore, M. Vievra, B. Timmerman, D. Feldon, and M. Maher, “The relationship between undergraduate research participation and subsequent research performance of early career STEM graduate students,” *J. High. Educ.*, vol. 86, no. 6, pp. 834-863, Nov. 2015, doi: 10.1080/00221546.2015.11777386.

[26] I. Junichi, “Effect of self-generated questioning in a college lecture class on learners’ ability to ask questions,” *Bulletion of University of Teacher Education Fukuoka*, vol. 67, no. 4, pp. 46-55, 2018. Accessed: May, 23, 2024. [Online], Available: http://purl.org/coar/access_right/c_abf2

[27] T. Kariya, *The Principles of Compound Eye Though: The Creative PowerSswitch That Everyone Has*. Tokyo, Japan: Kou Dan Sya (in Japanese), 2002.

[28] D. Miller, R. Linn, and N. Gronlund, *Measuremen and Assessment in Teaching*, 11th ed. London, UK: Pearson, 2012.

[29] R. Reeder, T. Stein, and M. V. Peelen, “Perceptual expertise improves category detection in natural scenes”, *Psychonomic Bull. Rev.*, vol. 23, pp. 172-179, Jun. 2015, doi: 10.3758/s13423-015-0872-x.

[30] L. Busetto, W. Wich, and C. Gumdinger, “How to use and assess qualitative research methods,” *Neurol. Res. Pract.*, vol. 2, no. 14, May. 2020, doi: 10.1186/s42466-020-00059-z.

[31] S. Margaei, F. A. Yazdi, and M. Chandrashekara, “MAXQDA and its application to LIS research,” *Education, Computer Science,* Mar. 2019. Accessed: May, 23, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:155504878

[32] R. Delgado, and X. A. Tibau, “Why Cohen’s Kappa should be avoided as performance measure in classification,” *PLoS One*, vol. 14, no. 9, Sep. 2019, doi: 10.1371/journal.pone.0222916.

[33] T. Saito, “*学校教育における定期テストに関する研究の動向*,”(in Japanese), Annual report of the Education Network Center, Graduate School of Education, Tohoku University, Miyagi, JP, Rep. vol. 16, 2016. Accessed: May, 6, 2024. [Online]. Available: https://www2.sed.tohoku.ac.jp/~edunet/annual_report/2016/16-05_saito.pdf

[34] L. Nahar, “The effects of standardized tests on incorporating 21st century skills in science classrooms,” *In. Sci. Ed. J*, vol. 4, no. 2, pp. 36-42, May. 2023, doi: 10.37251/isej.v4i2.324.