Brain, Digital, & Learning

pISSN : 2384-2474
eISSN : 2586-7490

RESEARCH ARTICLE

Analyses of Question Types and Evidence Evaluation Types in Big Data Analysis for 7th Grade Students

Da-yeon Yoo[1], Soo-min Lim[2], Youngshin Kim[1*]
[1]Kyungpook National University
[2]Science Education Research Institute at Kyungpook National University

중학교 1학년 학생의 빅데이터 분석을 통해 생성된 의문 유형, 증거 평가 수준에 대한 분석

유다연[1] · 임수민[2] · 김영신[1*]
[1]경북대학교 · [2]경북대학교 과학교육연구소

*Corresponding Author: Youngshin Kim, kys5912@knu.ac.kr

ABSTRACT

The purpose of this study was to analyze the types of questions generated through big data analysis, establish hypotheses based on the generated questions, and determine authenticity by analyzing the types of evidence evaluation. To this end, 'COVID-19' and 'yellow dust', which are relatively familiar subjects and various types of big data have been accumulated in large quantities, were selected as research themes. The subjects of the study were 137 7th grade students, 72 male students, and 65 female students. The results of the study are as follows. First, in big data analysis, questions were created centered on conjectural questions. Second, among the types of evidence evaluation, the rejection type did not appear. Therefore, it is necessary to establish a teaching strategy so that various types of questions can be presented. In addition, it is necessary to teach how to critically judge collected big data and how to use various big data.

Key words: Scientific questions, evidence evaluation, big data analysis, middle school students

Check for updates

OPEN ACCESS

Brain, Digital, & Learning
2023, Vol. 13, No. 4, 383-395

https://doi.org/10.31216/BDL.20230024

Received: October 10, 2023
Revised: November 04, 2023
Accepted: November 08, 2023

Introduction

Due to the influence of the Fourth Industrial Revolution, attempts to create new value through the convergence of various information and technologies are increasing across all sectors of society, and as a result, the amount of data is increasing exponentially. Accordingly, interest in utilizing big data is increasing.

Big data refers to a set of data so large that it cannot be stored, managed, and analyzed using existing computing technologies, as well as related technologies and personnel (McKinsey et al., 2011). Big data has the characteristics of real-time or immediacy, such as large volume, various types of data such as structured, semi-structured, and unstructured, and very fast data creation and movement speed (Amir & Haider, 2015). Such big data has limitations in processing with existing information processing systems, so big data analysis technology is needed to generate meaningful information and knowledge through collection, processing, and analysis (Bok & Yu, 2017).

Interest in big data analysis is no exception in education (Kim, 2017; Lee et al., 2022). If big data analysis is used in education, the horizons of educational research can be expanded by exploring new variables, relationships between variables, policies, and public opinion that have been overlooked in existing theories (Yoo, 2016). And it can strengthen learning capabilities needed in future society such as creativity, personality, and convergence, analysis, prediction, and technology (Hwang, 2021). In addition, learners can use big data analysis techniques in their own lives to solve life problems. In this way, if big data analysis is introduced into education, it could be an opportunity to achieve a data-based educational revolution (Choi et al., 2012).

In addition, in the 2015 revised science curriculum, knowledge information processing competency, which can process knowledge and information in various areas, was selected as one of the core competencies (MOE, 2015). The 2022 revised science curriculum also focuses on cultivating knowledge and information processing capabilities. To this end, the goal is to train people with digital literacy and scientific literacy based on advanced science and technology (MOE, 2022).

Accordingly, research related to big data analysis is currently actively underway in the field of domestic science education, including the use of data analysis inquiry models (Son & Jeong, 2020), the use of data visualization tools (Lee, 2019), and research on the use of Python (Kim et al., 2019), and data-based scientific writing research (Park & Son, 2020). However, in these studies, no studies were conducted in which students directly used and analyzed big data in the process of scientific inquiry. Accordingly, there is a need to analyze the processes that students actually use in the process of scientific inquiry.

In order to conduct scientific inquiry, the generation of questions (Chin & Brown, 2002; Chin & Malhorta, 2002) and the evaluation and creation of knowledge are considered important (Minstrell & van Zee, 2000; NRC, 2000). Questions refers to the curiosity or desire to know that arises when observing natural phenomena and recognizing unstable problems, doubts, and uncertainties that cannot be explained with current knowledge (Lawson, 1995). Generating questions can be said to be a very important process in science learning that stimulates students' motivation to learn and leads them to active inquiry (Kim et al., 1999; Kwon et al., 2003; NRC, 2000). These scientific questions are classified into conjectural, causal, predictive, methodological and applied questions (Lee et al., 2004).

In addition, evidence evaluation, which occurs in the process of comparing expectations derived from hypotheses and results obtained from experiments (Kuhn & Pearsall, 2000; Yang et al., 2004), is also a very important element in the scientific inquiry process (Zimmerman, 2007). Evidence evaluation activities can cultivate students' thinking skills and provide opportunities for class improvement (Chinn & Brewer, 2001). This type of evidence evaluation is divided into four types: conditional acceptance, unconditional acceptance, rejection, and abeyance, depending on whether the experimental results are trustworthy and whether an explanation system has been established (Yang et al., 2004). Accordingly, there is a need to analyze the process of generating questions and evaluating evidence through big data analysis. In particular, the first year of middle school, the grade in which the importance of the integrated inquiry process begins to expand, can provide more important implications.

Meanwhile, most information agencies and research institutes have recently opened up big data to general researchers, allowing them to use big data for various research. Among them, it is not difficult to apply to students, and it is suitable for themes that are frequently used in daily life. Accordingly, we conducted an activity to analyze big data on the themes of 'COVID-19' and 'yellow dust', which are two themes that are affecting our daily lives in many ways.

Therefore, the purpose of this study is to analyze the types of questions that 7th grade students generate through big data analysis, and to analyze the types of evidence evaluations that establish hypotheses based on the generated questions and determine truth or falsity. In addition, gender differences in the question generation and evidence evaluation process were additionally analyzed. The results of this study are expected to provide educational implications in the context of scientific inquiry activities on how big data can be used in school science education.

Research Methodology

Participants

The participants of this study were 155 7th grade students in metropolitan D. Excluding 18 students who did not respond to question generation and evidence evaluation, the analysis was ultimately conducted on 137 students. Of these, 72 were male and 65 were female. The composition of the group was based on 3 students, and some groups consisted of 3 or more students. The group was composed of heterogeneous groups with different levels of achievement, and included at least one member of the opposite sex.

Themes

The themes of this study were 'COVID-19' and 'yellow dust', chosen by students to analyze the types of scientific questions generated based on big data and the types of evidence evaluations that support them. These themes contain a large amount of various forms of big data accumulated. The big data used in the study is reliable because it comes from national institutions, and at the same time, it is in a diverse and unstructured form. In addition, two different themes were selected to prevent the preceding task from influencing the subsequent task.

The theme of 'COVID-19' has accumulated various data at the national level related to the number of confirmed cases by date and medical data derived from it since the outbreak of the coronavirus. Big data related to the theme of 'COVID-19' can be found at: Coronavirus Disease-19 (https://ncov.kdca.go.kr/), Korea Disease Control and Prevention Agency (https://www.kdca.go.kr), and National Statistics Portal (https: Data published on //kosis.kr) was used. Students were first provided with data on the number of confirmed cases by day. Afterward, they were allowed to freely search and explore big data related to COVID-19 from the Korea Disease Control and Prevention Agency and the National Statistics Portal.

On the theme of 'yellow dust', observed values of yellow dust concentration by region have been accumulated at the national level over a long period of time. Big data related to yellow dust concentration was first provided to students as

Excel data presented by the Korea Meteorological Administration's Weather Nuri (http://www.weather.go.kr). Afterward, students were allowed to freely explore dates and regions related to yellow dust.

Inquiry Activity Sheet

The inquiry activity sheet was designed to collect data on the scientific questions students generate during the process of analyzing big data and the evidence evaluation process that supports them. Two types of inquiry activity sheet were developed each on the themes of 'COVID-19' and 'yellow dust.' The inquiry activity sheet was divided into the stages of question generation, data conversion, and conclusion drawing by revising and supplementing the contents of the question record presented in previous study (Lee et al., 2004) that classified question types. The developed inquiry activity sheet was open-ended, allowing students to freely record it during their activities. The inquiry activity sheet was attached in the appendix.

In the question generation stage, various materials related to the theme of 'COVID-19' and 'yellow dust' were examined and problems or questions were recorded. Among the recorded contents, one question related to the personal or social level was selected and described along with the reason. In the data conversion stage, data were organized from related websites and graphs were presented to explain the hypothesis. Afterwards, the independent and dependent variables were explained in the presented graph. In the conclusion stage, the cause and effect of the presented data were analyzed and presented as evidence. Lastly, it was designed to suggest individual or social response methods in relation to the data.

Analysis Framework

The analysis framework for analyzing the questions and evidence evaluation types generated in the process of analyzing big data is as follows. First, scientific questions are questions that arise from observing natural phenomena or attempts to find out. Accordingly, the type of scientific question used in previous study (Lee et al., 2004) was used as a classification framework. Question types were classified into conjectural questions, causal questions, predictive questions, methodological questions, and applied questions (Table 1).

Specifically, conjectural questions refer to questions that raise curiosity about the name, composition, structure, or function of an object, while causal questions refer to questions about the cause of a phenomenon. Predictive questions refer to questions about the phenomenon that appears when variables are changed, and methodological questions refer to questions about a new method to solve the problem in a different way. An applied question is a form of question that asks how observations can be used in everyday life. Lastly, all questions except methodological questions were sub-categorized into exploratory questions and verificational questions depending on whether or not the subject of the question could be specified.

Evidence evaluation refers to the process of judging the truth or falsity of a hypothesis based on the data explored. The type of evidence evaluation was described in Yang, Chae & Cho (2004) used the classification framework(Table 2).

Evidence evaluation types were divided into four types: conditional acceptance, unconditional acceptance, rejection, and withholding. Conditional acceptance is a case of constructing a reason or a new explanation for trusting the experimental results, and unconditional acceptance is a case of believing in the experimental results themselves. Rejection is a reaction in which one does not evaluate the results of an experiment as reliable and judges that one's predictions are correct while withholding refers to the inability to determine whether or not the results of the experiment are trustworthy.

Table 1. A framework for classifying scientific questions on the topic of yellow dust

Question Type	Sub-Type	Criteria
Conjectural	Exploratory	- What are the advantages and disadvantages of yellow dust? - What does yellow dust consist of?- What are the similarities and differences between fine dust and yellow dust?
	Verificational	Are there any benefits to yellow dust? Does the impact of yellow dust vary depending on age or region? Is yellow dust the worst in spring?
Causal	Exploratory	- What causes yellow dust? - Why is yellow dust harmful to the human body? - Why can yellow dust travel long distances?
	Verificational	Does yellow dust cause respiratory diseases? Is the cause of yellow dust due to sand or soil? If yellow dust occurs, will fine dust also occur?
Predictive	Exploratory	What changes in the source of yellow dust could affect the yellow dust concentration in Korea? What is the relationship between various meteorological phenomena and yellow dust?
	Verificational	As the yellow dust gets worse, will the incidence of respiratory diseases increase? If the factory operation rate at the source of yellow dust decreases, will the yellow dust concentration also decrease? If global warming is alleviated, will the yellow dust phenomenon also be alleviated?
Methodological		- What masks are available to prevent yellow dust? - How can we reduce yellow dust concentration?
Applied	Exploratory	- If fatty foods help with yellow dust, what other foods can help with yellow dust?
	Verificational	- If fatty foods like pork belly help prevent diseases caused by yellow dust, would eating butter that is high in fat help? - If an air purifier reduces the concentration of indoor fine dust caused by yellow dust, can charcoal, which has a porous structure, also be used?

Table 2. A framework for evidence evaluation

Evidence Evaluation Type	Criteria
Conditional Acceptance	After presenting table or graph data related to the hypothesis, the dependent variable appeared as the independent variable changed. What is the reason why this phenomenon occurs?
Unconditional Acceptance	The dependent variable simply changed as the independent variable in the table or graph data related to the hypothesis changed.
Rejection	After presenting table or graph data related to the hypothesis, the data is wrong for some reason.
Abeyance	- No evidence was found to support the hypothesis.

Class Design

For the study, classes were first conducted with teachers and students who agreed to the purpose and method of the study. There were a total of two teachers, one male and one female, who conducted the class, and all of them had more than 10 years of experience working in secondary schools. Classes were conducted in three sessions for each theme.

First, in the first session of the class, big data example data related to the theme were presented, and then students were asked to record all the scientific questions they generated through the data on their worksheets. The questions generated by students went through the process of being refined into a question form in order to be classified into question types. In addition, the teacher encouraged students to freely explore not only the big data example materials presented by the teacher, but also various big data materials not presented by the teacher. The teacher guided the students so that they could recognize the meaning of big data by accessing as many different types of data as possible.

In the second session, one of the questions generated in the first session was selected and guidance was given to conduct scientific investigation. Students were asked to set hypotheses according to the format of independent and dependent variables, and were allowed to freely explore tables and graphs that serve as evidence evaluation materials to prove the established hypotheses. When students found data that they believed supported the established hypothesis, they were asked to transfer it to their workbook.

In the third session, students were asked to explain in detail why the data transferred to the worksheet supported the hypothesis using terms such as independent variable and dependent variable. In addition, in order to solve the problems created, students had time to think about and present what solutions they could come up with based on the data they found.

Data Collection and Data Analysis

Data collection was conducted through inquiry activity sheet written by students during the big data analysis process. Students individually filled out inquiry activity sheet and were allowed to freely record questions and evidence evaluations during big data analysis activities.

The scientific questions generated by students were classified based on the question type classification framework. First, the sentences generated by students were cleaned to classify the types of scientific questions. Declarative sentences were converted to interrogative sentences and spelling was corrected. In the process of revising declarative sentences into interrogative sentences, sentences where the purpose of the interrogative was unclear went through the process of being specified through interviews with students. Lastly, in order to distinguish question types more carefully, students' expressions were refined and the question types were classified.

The type of evidence evaluation analyzed whether relevant tables or graphs that could solve the questions raised by students were correctly presented, and whether the presented content was correctly interpreted based on the dependent and independent variables. During this process, cases where a scientific question was not selected or related tables and graphs were not presented were excluded from the analysis. In cases where it was unclear whether tables or graphs among the data presented by students were interpreted correctly, the type of evidence evaluation was classified after confirming whether the data presented was correctly understood through interviews with the students.

After analyzing the types of questions and evidence evaluation types through the above process, a consultation process was held with two science education experts. In this process, parts analyzed differently between the researcher and the two science education experts were decided upon after consultation.

Afterwards, statistical verification was performed using the SPSS 26.0 program. Differences according to type of question and level of evidence evaluation and differences according to gender were verified.

Results

Scientific Questions

Scientific questions are divided into conjectural, causal, predictive, methodological, and applied types. At this time, except for methodological questions, the remaining question types are classified into two sub-types, exploratory questions and verificational questions, depending on whether or not the subject of the question can be specified. Table 3 shows the frequency of scientific question types generated by students during the big data analysis process on the themes of ‘COVID-19’ and ‘yellow dust.’

Table 3. Frequency of types of scientific questions by theme

n(%)

Theme \ Question type		Conjectural	Causal	Predictive	Methodo-logical	Applied	Total
COVID-19	Exploratory	123(37.1)	102(30.8)	12(3.6)	42(12.7)	0(0.0)	331(100.0)
	Verificational	19(5.7)	3(0.9)	30(9.0)		0(0.0)	
	Sub-total	142(42.9)	105(31.7)	42(12.7)	42(12.7)	0(0.0)	
Yellow dust	Exploratory	129(42.9)	61(20.3)	6(1.9)	56(18.6)	0(0.0)	301(100.0)
	Verificational	17(5.6)	11(3.7)	20(6.6)		0(0.0)	
	Sub-total	146(48.5)	72(24.3)	26(8.6)	56(18.6)	0(0.0)	

The types of questions formed by students in the process of analyzing the theme of 'COVID-19' were 142 conjectural questions (42.9%), 105 causal questions (31.7%), and 42 predictive and methodological questions (12.7%), and no student generated an applicable question. In conjectural and causal questions, there were relatively more exploratory questions than verificational questions, but in predictive questions, there were more verificational questions.

In the process of analyzing the 'yellow dust' theme, the most common types of questions generated by students were 146 conjectural questions (48.5%), 72 causal questions (24.3%) and 56 methodological questions(18.6%), followed by 26 predictive questions (8.6%). As for the 'yellow dust' theme, no applicable questions were raised. Among the types of questions, there were relatively more exploratory questions in conjectural and causal questions, but there were more confirmatory questions in predictive questions. Regardless of the theme, not a single applied question was created, and in conjectural, causal, and methodological questions, there were relatively many exploratory questions that could not specify the object of question, but in predictive questions, the object of question could be specified. More confirmatory questions were appearing.

Meanwhile, we analyzed whether there was a statistical difference in frequency according to the types of questions generated by students by theme (Table 4). On the theme of 'COVID-19', students were generating statistically significantly more conjectural question types than predictive, methodological, and applied question types ($p<.05$). Causal question types were generated statistically significantly more frequently than predictive, methodological, and application question types ($p<.05$). Predictive question types were statistically generated more frequently than methodological and applied questions ($p<.05$). And the methodological question type was statistically generated more frequently than the applied question type ($p<.05$).

Regarding the theme of 'yellow dust', students generated the most conjectural questions statistically significantly more than the rest of the question types ($p<.05$). Causal question types were generated more frequently than predictive, methodological, and applied question types, and predictive question types were generated statistically significantly more frequently than methodological and applied question types ($p<.05$). And the methodological question type was statistically generated more frequently than the applied question type ($p<.05$).

Table 4. Difference verification results by theme and question type

n(%)

Theme \ Question type	Conjectural	Causal	Predictive	Methodo-logical	Applied	x^2
COVID-19	$142(22.5)_{aef}^{**}$	$105(16.6)_{cdf}$	$42(6.6)_{acf}$	$42(6.6)_{def}$	0(0.0)	47.215^{*}
Yellow dust	$146(23.1)_{abc}$	73(11.6)ab	$26(4.1)_{abd}$	$56(8.9)_{bcd}$	0(0.0)	50.652^{*}
Total	288(45.6)	178(28.2)	68(10.7)	98(15.5)	0(0.0)	92.903^{*}

$^{*}p<.05$ ** The same letters mean that there is a statistical difference in post hoc testing.

Through this, it was confirmed that various types of questions were being generated in both the ‘COVID-19’ and ‘yellow dust’ themes. However, not all applicable question types were created. The applied question type is a question about how it can be used in everyday life and requires a high level of cognition. Accordingly, similar to previous research (Lim & Kim, 2015) that found that many students had difficulty in generating applicable questions, this study also found that they had difficulty in generating applicable questions.

The frequency of conjectural question types was found to be the highest in both the ‘COVID-19’ and ‘yellow dust’ themes. The fact that the percentage of conjectural question types is the highest is a different result from the study by Eom & Lee (2015), which analyzed question types among elementary school students. This is because questions were created through observation activities about scientific phenomena, and this study is believed to be due to questions being generated from unstructured data, that is, differences in the way data are provided.

Additionally, this study was conducted by Lee et al. (2004), the rate of methodological questions was higher than that of the study. We hope that research will be conducted on the reasons for this difference in research analyzing big data compared to previous research that generates questions in the process of scientific inquiry.

Evidence Evaluation

For big data analysis, the students' evidence evaluation types were analyzed in the process of interpreting the questions they selected and the data related to them. Table 5 shows the differences in the frequency of evidence evaluation types according to the two presented themes.

Table 5. Differences in frequency by evidence evaluation types by theme n(%)

Theme \ Evidence evaluation type	Conditional acceptance	Unconditional acceptance	Rejection	Withholding	Total	x^2
COVID-19	12(36.4)	9(27.3)	1(3.0)	11(33.3)	33(100.0)	6.802
Yellow dust	$20(71.4)_{abc}$**	$4(14.3)_{a}$	$0(0.0)_{b}$	$4(14.3)_{c}$	28(100.0)	14.896*
Total	$32(52.4)_{abc}$	$13(21.3)_{a}$	$1(3.2)_{b}$	$15(24.5)_{c}$	61(100.0)	18.764*

*p<.05 ** The same letters mean that there is a statistical difference in post hoc testing.

In the evidence evaluation on the theme of 'COVID-19', conditional acceptance was the most frequent at 12 times (36.4%), followed by withholding 11 times (33.3%), unconditional acceptance 9 times (27.3%), and rejection 1 time (3.0%). However, there was no statistical difference in the frequency of each type of evidence evaluation (p>.05). However, in the evaluation of evidence on the theme of ‘yellow dust’, conditional acceptance was the most common at 20 times (71.4%), followed by unconditional acceptance and withholding at 4 times (14.3%). At this time, among the types of evidence evaluation, conditional acceptance had a statistically significantly higher frequency than unconditional acceptance, rejection, and withholding (p<.05).

Looking at the evaluation of evidence across themes, conditional acceptance was the highest at 32 times (52.4%), followed by withholding at 15 times (24.5%) and unconditional acceptance at 13 times (21.3%). At this time, conditional acceptance was statistically significantly higher than unconditional acceptance, rejection, and withholding (p<.05).

The fact that conditional acceptance has the highest frequency among evidence evaluations means that students can find evidence to explain the questions they have generated and have the ability to construct an explanatory system related to it. Meanwhile, unconditional acceptance also accounted for a certain portion, indicating that students recognized that the contents of tables and graphs they found were related to solving questions, but accepted them unconditionally because they could not provide scientific explanations. Through this, half of the middle school students were able to find evidence for questions and construct an explanation system, and it was confirmed that the remaining students were required to develop the ability to provide scientific explanations.

The frequency of rejection due to lack of trust in the experimental results was the lowest. Comparing this to previous research (Witzig at al., 2013), which found that students evaluate texts that contain a lot of factual information or are written by authoritative authors as more trustworthy information, it is not the data obtained by students conducting their own experiments, but rather the data obtained by conducting experiments on their own. It can be inferred that in the process of searching for data officially collected by public institutions, results supporting the hypothesis were found based on trust in clear sources.

Approximately 25% of respondents expressed reservations about evaluating evidence. The main reason these students identified was not being able to find data to support their questions. Accordingly, a teaching and learning strategy that provides additional relevant materials to solve the questions raised by students is needed.

Evidence Evaluation according to Scientific Questions

The results of analyzing how the type of evidence evaluation varies depending on the selected question are as follows. Table 6 is the result of verifying the frequency and differences in evidence evaluation types for each type of question selected by students.

Among the 27 types of conjectural questions, conditional acceptance was found 14 times (51.9%), withholding was found 9 times (33.3%), and unconditional acceptance was found 14 times (14.8%). In the causal question type, the evidence evaluation type of conditional acceptance was the most common with 9 cases (47.4%), followed by unconditional acceptance and withholding with 5 cases each (26.3%). The predictive questioning type also showed similar results to the other types, with conditional acceptance was found 10 times (66.6%), unconditional acceptance was found 3 times (20.0%), and rejection and withholding was found once each (6.7%). As shown above, the rate of conditional acceptance, which is the type that trusts the experimental results and forms an explanation system for the results, was the highest among all question types, and the frequency of the rejection type was very low regardless of the question type.

Table 6. Verification of difference in frequency in evidence evaluation by scientific questions n(%)

Question type \ Evidence evaluation type	Conditional acceptance	Unconditional acceptance	Rejection	Withholding	Total	x^2
Conjectural	$14(51.9)_{ab}$**	$4(14.8)_{a}$	$0(0.0)_{b}$	9(33.3)	27(100.0)	7.543*
Causal	$9(47.4)_{abc}$	$5(26.3)_{a}$	$0(0.0)_{b}$	$5(26.3)_{c}$	19(100.0)	5.012*
Predictive	$10(66.6)_{abc}$	$3(20.0)_{a}$	$1(6.7)_{b}$	$1(6.7)_{c}$	15(100.0)	4.711*
Total	$33(54.1)_{abc}$	$12(19.7)_{a}$	$1(1.6)_{b}$	$15(24.6)_{c}$	61(100.0)	15.374*

*p<.05 ** The same letters mean that there is a statistical difference in post hoc testing.

As a result of analyzing the statistical differences in evidence evaluation types according to question type, it was found that conditional acceptance was statistically more frequent than unconditional acceptance, rejection, and withholding evidence evaluation types in all question types ($p<.05$). However, in the conjectural question type, there was no statistically significant difference between conditional acceptance and withholding.

Discussion

The purpose of this study is to analyze the types of questions that 7th grade students generate in the process of analyzing big data and the level of evidence evaluation accordingly. The discussion of this study is as follows. First, the proportion of scientific questions generated by 7th grade students based on big data was conjectural, followed by causal, methodological, predictive, and applied questions. This is consistent with research showing that 75% of students' questions were in the form of factual questions (Marbach-Ad & Sokolove, 2000; Stano, 1981), However, it was inconsistent with a previous study targeting elementary school students and pre-service elementary school teachers (Lee et al., 2004) and a previous study on scientific questions generated targeting middle school students (Lim & Kim, 2015). In previous studies (Lee et al., 2004; Lim & Kim, 2015), causal questions appeared at the highest rate, followed by conjectural, predictive, applied, and methodological questions. This difference is believed to be due to the fact that, unlike previous studies, the form of data presented in this study was presented in various concepts, and quantified values related to it were presented.

In addition, the overall higher frequency of exploratory questions than verificational questions is believed to be because the participants of this study were 7th grade students and did not have much prior knowledge related to the theme (Lee et al., 2004). Therefore, before implementing science teaching-learning using big data, it is necessary to take steps to activate students' prior knowledge before class.

The type of evaluation of questionable evidence generated based on big data is also described by Yang, Chae & Cho (2004) was consistent with the study. The type of evidence evaluation made by 7th grade students based on big data was the highest type of conditional acceptance, followed by withholding, unconditional acceptance, and rejection. This shows that students evaluate evidence based on trust in the sources of graphs and tables needed to explain scientific questions created based on big data. This result is consistent with Witzig at al. (2013) showing that students evaluate texts that contain more factual information or are written by authoritative authors with greater confidence. It can be thought that the reason why the frequency of the withholding type accounts for a certain amount is because students were unable to find the material they wanted among a large amount of information. Therefore, it is suggested that there is a need to provide information on students' critical judgment ability and data search ability in the process of using big data in actual classes.

The direction of scientific inquiry activities is determined by the type of question (Lee et al., 2004). However, the fact that no applicable questions are raised for each theme in the results of this study means that various forms of inquiry activities are not carried out. Accordingly, in order to generate various types of questions, there is a need to analyze the properties of the task in advance and present various types of big data before the teaching-learning process. In order to change the evaluation type from evidence evaluation activities to conditional acceptance, various teaching-learning strategies must be used. For example, alternative concepts that correspond to new concepts can be presented (Noh et al., 2000) or cognitive conflict can be induced and replaced with the correct scientific concept and then instructed to evaluate the evidence (Kwon & Kwon, 2004). Additionally, methods to improve search capabilities, such as presenting search-related words in the dictionary, could be applied.

Students cannot find answers to questions right away, and it is difficult to directly connect questions to research (Chin & Kayalvizhi, 2002; Garlick & Laugksch, 2008). Therefore, teaching and learning that can generate researchable questions that enable students to continuously accumulate experience should be provided by providing various quantitative big data. Additionally, because anomalous cases result in different response patterns from students (Chinn & Brewer, 1993), it would be more effective to provide data that could result in increases that do not match expectations.

Conclusion

This study analyzed the types of scientific questions generated and the types of evidence evaluated to support them by 7th grade students based on big data. The themes were 'COVID-19' and 'yellow dust', for which sufficient big data has been accumulated. Based on the collected data, this study analyzed types of questions and types of evidence evaluation. The conclusions of this study are as follows.

First, the generation of questions in big data analysis focuses on some types of questions. The majority of questions generated based on big data were conjectural questions, and many of them were exploratory questions. Although the types of questions generated most frequently were different from previous studies related to question types, the number of methodological or application questions generated was small in both the previous studies and this study, and the focus was on some question types. Therefore, there is a need to establish a teaching strategy so that various types of questions can be presented.

Second, some types of evidence evaluation did not present. It was encouraging to see that the rate of conditional acceptance was high. However, the fact that students trusted the data they searched for related to the theme of inquiry and were able to easily design an explanation system has significant implications regarding critical thinking and the difficulty of the task. First, students should be instructed to critically judge the data they search for and then decide whether or not to accept the data. In addition, big data analysis should be conducted on themes with an appropriate level of difficulty to encourage students' sense of challenge.

Based on the results of this study, it is necessary to provide a wide range of big data that can also generate applicable questions through big data analysis. In addition, a teaching strategy that guides students to generate applicable questions in problem situations related to big data is required. In addition, in big data analysis, analysis of how research activities are conducted according to the type of question and analysis of various forms of big data will be conducted.

Conflicts of Interest

The authors declared no conflicts of interest.

References

Amir, G., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35, 137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Bok, K., & Yu, J. (2017). Big Data in the Fourth Industrial Revolution. Journal of KIISE, 35, 29-39.

Chin, C. A. & Malhorta, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86. 175-218.

Chin, C., & Brown, E. D. (2002). Posing problems for open investigations : What questions do pupils ask?. Research in Science & Technological Education, 20, 269-287.

Chin, C., & Kayalvizhi, G. (2002). Posing problems for open investigations: What questions do pupils ask? Research in Science & Technological Education, 20, 269-288.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1-49

Chinn, C. A., & Brewer, W. F. (2001). Models of data : A theory of how people evaluate data. Cognition and Instruction, 19, 323-393.

Choi, J., Park, C., Choi, K., Jeong, E., Kim, S., & Yoo, I. (2012). Application of educational big data generated in smart education. Korea Intelligent Information System Society, 2012, 144-148.

Garlick, R., & Laugksch, R. C. (2008). Teaching children to ask investigable questions in science. School Science Review, 90, 119-127.

Hwang, H. (2021). A case study on application of big data-based social studies teaching and learning model. Social studies education, 60, 111-131.

Kim, D. (2017). Identity and role of elementary education in the fourth industrial revolution era. The Journal of Korean Educational Idea. 31, 23-45. https://doi.org/10.17283/jkedi.2017.31.4.23

Kim, J., Kim, M., Yoo, H., Kim, Y., & Kim, J. (2019). Effect of data visualization education with using Python on computational thinking of six grade in elementary. Journal of The Korean Association of Information Education, 23, 197-206.

Kim, J., Kim, M., Yu, H., Kim, Y., & Kim, J. (1999). Effect of data visualization education with using Python on computational thinking of six grade in elementary. Journal of The Korean Association of Information Education, 23, 197-206.

Kwon, N., & Kwon, J. (2004). The effects by learners' characteristics on scientific conceptual changes using cognitive conflict strategy. Journal of The Korean Association For Science Education, 24, 216-225.

Kwon, Y., Jeong, J., Kang, M., & Kim, Y. (2003). A grounded theory on the process of generating hypothesis - knowledge about scientific episodes. Journal of The Korean Association For Science Education, 23, 458-469.

Lawson, A. E. (1995). Science Teaching and the Development of Thinking. Arizona: Wadsworth Publishing Company.

Lee, H. (2019). The effects of STEAM program utilizing data visualization tools on elementary students' ability to process knowledge information, attitudes towards science and technology (Unpublished Master's Thesis). Korea National University of Education, Cheongju, Republic of Korea.

Lee, H., Jeong, J., Park, K., & Kwon, Y. (2004). Types of scientific questions generated in observational activity by elementary students and preservice teachers. Journal of The Korean Association for Science Education, 24, 1018-1027.

Lee, J., Lim, Y., Bae, H., Jeon, H., Shin, C., & Noh, J. (2022). Exploring the Curriculum Implementation Support System for the 2022 Revised Curriculum (RRC 2022-3). Korea Institute for Curriculum and Evaluation.

Lim, S., & Kim, Y. (2015). Analysis of correlation between scientific questions and used knowledge types from middle school students. Brain, Digital, & Learning, 5, 15-24.
Marbach-Ad, G. & Sokolove, P. G. (2000). Can undergraduate biology students learn to ask higher level questions? Journal of Research inScience Teaching, 37. 854-870
McKinsey, J., Chui, M., Brown, B., Bughin,J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation. Retrieved October 24, 2022. from https://www.mckinsey.com/capabilities/mckinsey–digital /our-insights/big-data-the-next-frontier-for-innovation.
Ministry of Education [MOE] (2015). Science Curriculum (The notification of MOE 2015-74: The separate volume no. 9). Ministry of Education.
Ministry of Education [MOE] (2022). Science Curriculum (The notification of MOE 2022-33: The separate volume no. 9). Ministry of Education.
Minstrell, J., & van Zee, E. (Eds). (2000). Inquiring Into Inquiry Learning and Teaching in Science. Washington, D. C.: American Association for Advancement of Science.
National Research Council (2000). Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Academy Press.
Noh, T., Lim, H., & Kang, S. (2000). Types of students' responses to anomalous data. Journal of The Korean Association For Science Education, 20, 288-296.
Park, C., & Son, J. (2020). The effects of data-based scientific inquiry linked with Science Writing Heuristic (SWH) on elementary school students' science core competencies. Teacher Education Research, 59, 245-258.
Son, M., & Jeong, D. (2020). Development of data-driven science inquiry model and strategy for cultivating knowledge-information-processing competency. Journal of The Korean Association For Science Education, 40, 657-670.
Stano, A. S. (1981). A study of the relationship between a specific question-asking technique and students' achievement in higher cognitive level question asking (Unpublished Doctoral Dissertation). University of Chicago, Chicago, USA.
Witzig, S. B., Halverson, K. L., Siegel, M. A., & Freyermuth, S. K. (2013). The interface of opinion, understanding and evaluation while learning about a socioscientific issue. International Journal of Science Education, 35, 2483-2507.
Yang, I., Chae, K., & Cho, H. (2004). Analysis of thought processes occurring in scientific evidence evaluation activities. Korean Journal of Teacher Education, 21, 276-308.
Yoo, J. (2016). An analysis case of educational panel data through a data mining technique: A penalized regression with KYPS data. Asian Journal of Education, 17, 1-19.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. Development Review, 27, 172-223.

Authors Information

Yoo, Da-yeon: MS Student, Kyungpook National University, First Author.

Lim, Soo-min: Research Professor, Science Education Research Institute at Kyungpook National University, Co-author

Kim, Youngshin: Professor, Kyungpook National University, Corresponding Author