Brain, Digital, & Learning

pISSN : 2384-2474
eISSN : 2586-7490

RESEARCH ARTICLE

Development and Application of a Science Inquiry Instruction Program for Big Data Analysis: Focusing on COVID-19

Bo-kyung Kim[1], Youngshin Kim[2*]
[1]Hamji High School
[2]Kyungpook National University

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

ABSTRACT

The rapid advancement of technologies such as artificial intelligence, big data, and the Internet of Things has reshaped modern society into an environment where large volumes of information must be processed in real time. As a result, data literacy and information-processing competencies have become indispensable for students. Science education is well positioned to cultivate these abilities because its inquiry processes—observation, data collection, analysis, and interpretation—closely parallel the steps involved in data analysis. To address these limitations, this study developed a science inquiry program utilizing publicly available COVID-19 datasets, a topic closely linked to students' everyday experiences, and implemented it with 125 second-year high school students in D Metropolitan City. The three-hour instructional program comprised five stages: (1) understanding big data, (2) identifying problems, (3) planning data collection, (4) analyzing and visualizing data, and (5) drawing conclusions and presenting findings. The big data analysis activities were conducted individually, and students freely described their work in Google Slides submitted through Google Classroom. The effectiveness of the classes was evaluated through data literacy assessments and class satisfaction surveys. The findings show that the big-data-based inquiry activity using public data had a positive impact on students' data literacy and satisfaction with the class. Additionally, the results underscore the importance of continuous teacher guidance and feedback throughout the inquiry process. Based on these outcomes, this study highlights the need for future science instruction to incorporate not only quantitative but also qualitative approaches to big data analysis.

Keywords: Big data analysis, data literacy, COVID-19, class satisfaction, use of public data

Check for updates

OPEN ACCESS

Brain, Digital, & Learning
2025, Vol. 15, No. 4, 601-615
https://doi.org/10.31216/BDL.2025.15.4.5

Received: November 24, 2025
Revised: December 15, 2025
Accepted: December 15, 2025

Introduction

As cutting-edge technologies such as artificial intelligence, big data, and the Internet of Things rapidly diffuse throughout society, we now inhabit an environment in which vast quantities of data are encountered in real time (Kusuma & Viswanath, 2018; Popchev et al., 2019; Velarde, 2020). In the midst of these societal changes, school education can no longer simply deliver predetermined answers. Beyond memorization and the reproduction of information, students must increasingly develop the ability to identify problems, explore data-based solutions, and derive conclusions through rational interpretation and decision-making (Mandinach & Gummer, 2016; OECD, 2018). In modern society—marked by growing uncertainty and complexity—individuals are required to continually select information, compare and analyze diverse data, and engage in independent thinking and judgment. Consequently, competencies such as data literacy and knowledge-information processing—encompassing the ability to read and analyze data, communicate effectively, and collaborate based on data—are becoming essential for future citizens (Bae, 2019; MOE, 2022; Song et al., 2021).

To cultivate these competencies, educational practice is shifting from teacher-centered knowledge transmission toward student-centered problem solving, data-driven inquiry, and collaborative learning (MOE, 2022; OECD, 2018). Within this shift, science curricula must offer learning opportunities that allow students to analyze real-world social issues, interpret them scientifically, and experience rational decision-making through data-driven inquiry activities (D'Ignazio & Klein, 2020).

Science education is particularly well suited for fostering these emerging competencies (Son & Jeong, 2020). The scientific inquiry process—comprising observation, data collection, analysis, interpretation, and decision-making—mirrors the stages of data analysis and provides students with authentic opportunities to apply scientific thinking using real-world data. Rather than simply delivering fragmented knowledge, science serves as a tool for explaining and interpreting real phenomena, positioning it closely within the broader context of data-driven problem solving (Son & Jeong, 2020).

However, this pedagogical shift has not yet been fully realized in actual science classrooms. Existing big-data-based instructional research has often targeted specialized populations such as science high schools, gifted education programs, or university majors (Shin & Choi, 2021; Song & Choi, 2022), leaving few examples applicable to general high school settings. Moreover, many current studies assume the use of complex tools or programming languages or focus on abstract concepts, which can create barriers for general high school learners. Thus, there is a need for instructional designs that enable students to engage in accessible, relevant, data-exploration activities grounded in familiar real-world contexts.

To address this need, the present study developed and implemented a science inquiry program utilizing publicly available "COVID-19" big data—a topic closely connected to students' daily lives. COVID-19, as a social issue with direct personal relevance and biological dimensions, provides rich material for scientific inquiry and data analysis. Therefore, this study explored how science classes should be designed to support big data analysis. In addition, it examined the program's impact on students' data literacy and their overall satisfaction. Through this study, we aim to establish a foundation for teaching-and-learning design that supports big-data-based analysis within general science education.

Research Methodology

Subjects

This study was conducted with 181 second-year students at a high school in Metropolitan City D who provided informed consent after being informed of the study's purpose. The school's overall academic achievement level is in the upper–middle range nationally. After excluding 56 students who did not complete either the pre- and post-data literacy surveys or the class satisfaction survey, data from 125 students were ultimately included in the analysis.

Instruction Program

The course program consisted of five stages: understanding big data, identifying a problem, planning data collection, analyzing and visualizing big data, and drawing conclusions followed by presentations. The step-by-step process was organized as follows. First, in the Understanding Big Data phase, students examined examples of big data used in everyday life and learned how such data functions in society. They then explored different types of big data and were introduced to basic tools for data analysis.

In the Problem Recognition phase, students selected a topic to investigate using big data. To support this process, we provided a variety of COVID-19–related materials and guided students to identify a problem of interest. Based on the questions they formulated, students developed hypotheses and predicted possible outcomes before proceeding to the next stage of inquiry.

In the Data Collection Planning stage, students established a detailed plan for conducting big data analysis. After examining various examples of big data visualization, they selected COVID-19–related datasets appropriate for testing their previously formulated hypotheses. Based on the chosen datasets, students identified relevant variables and determined suitable analysis methods and visualization strategies.

In the Big Data Analysis and Visualization stage, students constructed graphical representations of the collected data and interpreted these visualizations to derive meaningful insights. In the Conclusion and Presentation stage, students

Table 1. Scientific exploration process for big data analysis

Steps	Activity
Understanding big data	• Explore examples of big data utilization • Explore various big data resources • Explore big data analysis tools
Problem recognition	• Big data exploration • Identifying problems or questions to be addressed • Hypothesis development
Data collection plan	• Collect data to support the hypothesis • Select big data variables and analysis techniques • Plan a visualization method
Big data analysis and visualization	• Visualize selected big data data • Title the graph • Interpret the research results
Draw conclusions and present	• Draw conclusions about the hypothesis based on the visualized data • Explain future tasks and implications • Organize and present in PowerPoint

drew conclusions related to their research questions using the analyzed and visualized data. They then prepared and presented reports that included the implications of their findings and suggestions for future inquiry.

The developed lesson program was evaluated by five experts: one researcher with a doctorate in science education and four teachers holding master's degrees. The expert review indicated a validity level of 92%, confirming the appropriateness of the program.

Student Activity Sheet

The student worksheets were structured according to the sequence of the instructional program: understanding big data, identifying the problem, planning data collection, analyzing and visualizing data, and drawing conclusions followed by presentations. The worksheets were distributed individually in Google Slides format via Google Classroom, and students were encouraged to respond freely to each item.

The specific components of the worksheets were as follows. In the Understanding Big Data stage, the worksheets presented the learning objectives and provided guidance on relevant online resources—such as the National Statistical Portal, Health and Welfare Data Portal, Statistics Data Center, Seoul Open Data Plaza, Public Data Portal, Environmental Data Platform, and Google Trends—to support students in exploring and understanding various types of data.

The Problem Recognition stage consisted of three steps: identifying questions, selecting an inquiry topic, and developing hypotheses. In the Identifying Questions step, students were encouraged to freely write about social, economic, cultural, and scientific changes since COVID-19, issues they personally wished to address, and any additional questions of interest. During the Selecting an Inquiry Topic step, students chose one of their previously formulated questions as their final research topic and described the background and relevant curriculum content that led to that question. In the Developing a Hypothesis step, students generated a hypothesis for the selected question.

The Data Collection Planning stage required students to design an analysis strategy to address their research topic. They were instructed to explicitly describe the data sources, variables to be analyzed, and appropriate analysis methods. In the Big Data Analysis and Visualization stage, students collected and visualized data according to their plan, interpreted the resulting graphs, and extracted meaningful insights.

Finally, in the Conclusion and Presentation stage, students drew conclusions based on the analyzed data and prepared a report that included the implications of their findings. They assessed data sources, data descriptions, collected items, patterns of change, and relationships among variables to evaluate the validity of their hypotheses. Based on this analysis, students identified future challenges, practical implications, real-life relevance, and potential solutions.

Questionnaires

Two questionnaires were developed: one assessing data literacy and one assessing class satisfaction. The data literacy test utilized a tool developed by Song et al. (2021). Data literacy refers to a set of activities involved in collecting, analyzing, and interpreting the various types of data encountered in everyday life (Swan et al., 2009). According to Bae (2019), data literacy consists of two main areas: statistical analysis and data utilization and communication. Statistical analysis encompasses understanding, analyzing, and visualizing data, whereas data utilization and communication involves using data to convey information and share insights with others.

Data literacy questionnaires consists of two areas: "Statistical Analysis" and "Data Utilization and Communication," each subdivided into three subareas, for a total of 21 questions. All questions were answered on a 5-point Likert scale.

Table 2 presents the structure of the data literacy test items and the reliability (Cronbach's α) for each subarea. The reliability of all data literacy items was very high, with a Cronbach's α of 0.959.

Table 2. Question structure and reliability of the data literacy test

Domains	Sub-domain	Question number	No. of questions	Reliability
Statistical analysis	Understanding Data	1~3	3	.765
	Data Collection and Preparation	4~7	4	.815
	Data Analysis	8~11	4	.887
Data-driven communication	Data Presentation	12~14	3	.802
	Data Evaluation	15~17	3	.844
	Data-Driven Communication	18~21	4	.902
Total		1~21	21	.959

The class satisfaction survey was reconstructed based on the class satisfaction questionnaire developed by Kim (2019) and Hwang (2018). The evaluation area consisted of three sub-areas: class objectives, class content, and class effectiveness. Each sub-area contained five questions, for a total of 15 questions. All questions were answered on a 5-point Likert scale. The overall reliability of the developed class satisfaction survey was Cronbach's α = 0.951, and the reliabilities of each sub-area ranged from 0.873 to 0.909.

Instruction Management

The classes were conducted by teachers who consented to the study's purpose and methods. A total of three teachers participated, with each school assigning a biology teacher to implement the program. The teachers voluntarily conducted the big data–based, inquiry-oriented instructional program developed in this study after receiving a thorough orientation to the program and its procedures. Prior to teaching, the teachers were provided with detailed guidance on the research design and instructional process. The participating teachers had 10 to 22 years of teaching experience and included one male and two female teachers. All had graduated from teacher training colleges, with qualifications ranging from a bachelor's degree to a master's degree, and one holding a doctorate.

The instructional program was delivered over three 50-minute sessions, with each teacher facilitating the learning process, coordinating the inquiry activities, and providing feedback. The class followed a five-stage structure: (1) learning guidance and understanding big data, (2) problem identification, (3) data collection planning, (4) big data analysis and visualization, and (5) drawing conclusions and implications.

The first session focused on understanding big data and identifying research problems. Students explored various big data portal sites to examine real-world datasets and, based on this exploration, formulated questions related to societal changes and personal interests in the context of COVID-19. They then selected research topics from these questions and developed corresponding hypotheses. Teachers provided individual feedback on the hypotheses via Google Classroom, and students reflected on and revised their hypotheses accordingly.

During the second session, students planned data collection and conducted big data analysis and visualization activities. Using the data sources provided by the teacher, students collected relevant data to test their hypotheses. They identified manipulated and dependent variables, organized raw data in Excel, and visualized the results. Completed graphs and analysis outputs were submitted individually through Google Classroom.

The third session centered on drawing conclusions and implications. Students evaluated the validity of their hypotheses based on the analyzed data, summarized their findings, and drew conclusions with practical and societal implications. At the end of the activity, students completed a reflective section on the worksheet, providing an opportunity to assess their learning and the overall inquiry process.

Data Collection and Analysis

To develop a program for big data analysis and verify its effectiveness, a treatment group was established. This group completed a data literacy assessment before the class, and after the class, they were evaluated on both data literacy and class satisfaction.

Data was collected through students' Big Data Exploration Worksheets and questionnaires. The student worksheets contained information recorded by students in Google Classroom during the class. They included questions related to COVID-19, reasons for selecting the research topic and related course content, hypotheses, data collection plans, data visualization results and interpretation, conclusions, and future challenges and implications.

The survey was conducted through data literacy and class satisfaction tests. The data literacy tests were administered before and after class, and the class satisfaction test was administered after the big data class. All surveys were supervised by the teacher in charge of each school. Students were allowed to freely respond to their own thoughts, and the surveys took approximately 10 minutes to complete.

Both the data literacy and class satisfaction tests used a 5-point Likert scale, with responses ranging from "not at all" (1 point) to “very much” (5 points). Therefore, the data literacy test scores ranged from a minimum of 21 to a maximum of 105, while the class satisfaction test scores ranged from a minimum of 14 to a maximum of 75. Student activity sheets were analyzed qualitatively through a qualitative analysis of student records. Data literacy questionnaires were analyzed using the statistical analysis program SPSS version 26. Paired-samples t-tests were used to calculate pre- and post-test means, standard deviations, and mean differences. Statistical significance was assessed at a 5% significance level. The class satisfaction survey calculated mean and standard deviations by domain to assess satisfaction with class objectives, course content, and instructional effectiveness.

Results and Discussion

Activity Process

Problem Recognition

In the problem identification stage, students explored social changes related to COVID-19, and then developed their own research topics based on this. During this process, students used the various examples of social change presented by the teacher as clues, focusing on recalling changes they experienced or questions they had in their daily lives. Furthermore, students freely recorded their experiences across various categories, including society, economy, education,

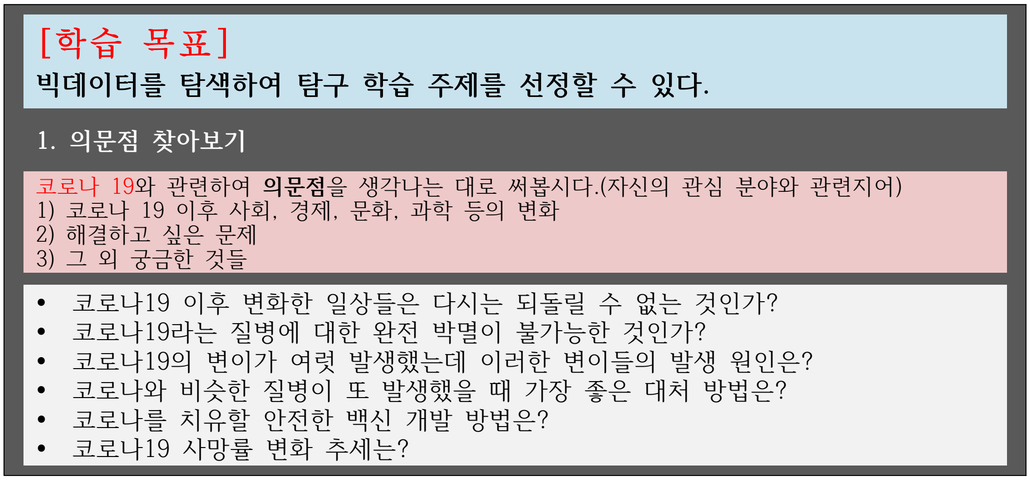

Fig. 1. Example of a student-written question-and-answer

environment, emotions, and health, based on their own interests. Fig. 1 shows an example of a student's activity during the question exploration stage.

The questions raised by students were somewhat related, while others presented them in fragmented, unrelated terms. Furthermore, the number of questions varied across students. Below are examples of questions written by students:

• To what extent have the environment and ecosystems seen positive changes due to reduced human intervention since COVID-19?

• With human activity intensifying again after the pandemic, how can we minimize environmental damage and maintain restored ecosystems?

• What environmental changes have occurred before, during, and after the COVID-19 pandemic?

• What impact has COVID-19 had on individual economies?

• What are some cases where COVID-19 patients are asymptomatic?

• How has participation in leisure activities changed during the COVID-19 pandemic?

• Has COVID-19 significantly impacted people with disabilities?

• Has COVID-19 had a significant impact on the healthcare industry?

• Changes in medical applications after COVID-19

• Changes in healthcare service utilization after COVID-19

• Which sectors have promoted contactless services and experienced increased sales since COVID-19?

• Will there be a significant increase in disposable product usage?

• Are people still happy when they spend long periods at home?

Students then selected a research topic based on both analyzability and interest. They explained the motivation and purpose of the research related to the selected question and developed curriculum content related to the topic.

Based on the selected topic, students developed a hypothesis that could be tested through data. Initially, the manipulated and dependent variables were often unclear or simply descriptive, and the hypotheses were difficult to verify with data. Therefore, the teacher provided feedback on the students' hypotheses through Google Classroom, guiding them to reflect more deeply and revise and improve their hypotheses. Fig. 2 provides examples of hypotheses developed by students in the initial stages.

2. 탐구 주제 선정하기

여러 의문점 중 한 가지를 최종 선정하여 그 의문점을 가지게 된 배경을 설명합니다.

탐구 주제 (의문점)	코로나19 이후의 실업률 상승
탐구 동기 및 목적 (의문점을 가지게 된 배경)	코로나19 이후 어떤 산업에서 실업률이 증가했고 어떤 산업에서 실업률이 감소했는지가 궁금하다.
주제 관련 교과 내용	경영향: 기업의 인재 관리와 예산 분배를 위해 취업과 해고를 조절한다.

3. 가설 설정하기

선정한 의문점에 대해 자기 나름의 가설을 설정합니다.

가설	의료, 보건 산업에서는 실업률이 감소하고 자영업에서는 실업률이 증가했을 것이다.

Fig. 2. Examples of student responses to selecting a research topic and setting a hypothesis

• As the number of COVID-19 cases increased and the economy gradually slowed, the demand and use of delivery food steadily increased. Based on this, demand is expected to continue to increase even after the COVID-19 pandemic is over.

• Developed countries are expected to have high initial vaccination rates, while overall, vaccination rates are expected to be low due to limited vaccine supply and side effects.

• Negative psychological changes are likely to have occurred following COVID-19.

• The shift to online, non-face-to-face systems following COVID-19 has likely increased productivity and learning effectiveness.

• Emotional distress may have increased during the COVID-19 pandemic.

• The number of delivery app users is expected to increase in 2020 and 2021.

• Since COVID-19, rising fuel costs and reduced supply have led to increased airfare prices, which may have led to decreased customer demand.

• Efficient design of nanoparticle delivery systems can enhance the immune response of vaccines, thereby reducing the required dose and improving distribution efficiency.

In this process, students distinguished between manipulated and dependent variables and examined whether their hypotheses could be analyzed using actual big data. They were asked to consider questions such as, "Can we find the relevant data?" and "How can we express changes numerically?"

Data Collection Plan

The data collection planning stage focused on specifying how and where to collect the big data necessary to verify the established hypotheses. Students focused on concretizing their research topics into a form that could be practically analyzed.

1. 데이터 수집

가설을 뒷받침 해 줄 데이터를 수집하여 그 출처를 링크로 첨부한다.

데이터 출처	https://kosis.kr/statHtml/statHtml.do?orgId=106&tblId=DT_106N_03_0200073&vw_cd=MT_ZTITLE&list_id=T_7&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE
활용할 변인	대기오염 정도 (대기 중 이산화 질소 농도) 코로나 전, 코로나 기간
데이터 출처	https://kosis.kr/statHtml/statHtml.do?orgId=355&tblId=TX_35501_A068&vw_cd=MT_ZTITLE&list_id=355_35501_9&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE
활용할 변인	멸종위기종 개체수
데이터 출처	https://www.k-stat.go.kr/metasvc/msba100/statsdcdta?statsConfmNo=106001&kosisYn=Y
활용할 변인	수질오염 정도

Fig. 3. Example of student responses to data collection plan

Students divided their hypotheses into observable variables—the manipulated and dependent variables—and based on this, they developed a detailed plan for the actual big data collection. Finally, students created links to data sources and the variables they would utilize within those data in a Google Slides presentation and submitted the raw data as an Excel file to Google Classroom. This stage was structured to concretize strategies so that students could selectively collect data according to the logical structure of their proposed problem, rather than simply "searching" for data.

At this stage, many students felt unfamiliar with converting variables linked to hypotheses into quantifiable data, leading to a sense of uncertainty. Furthermore, despite expectations of data availability, some students were unable to find the desired data on public platforms, forcing them to re-establish their data collection paths. Some students, instead of directly visualizing big data, simply collected visualizations provided on public platforms.

Therefore, teachers should provide individual feedback to students to assess the feasibility and direction of their data collection, and provide guidance on how to directly collect raw data from statistical websites.

Analysis and Visualization

Based on the collected big data, students conducted quantitative analyses. Each student selected an appropriate analysis method based on the characteristics of the selected analysis variables and the data structure. Excel was used as a visualization tool. Most of the data was provided in Excel format, making visualization relatively simple, so learning the method didn't take long. The resulting graphs were captured and inserted into Google Slides, and students were instructed to document the graph's meaning. Fig. 4 shows an example of data visualization.

Students used line graphs to analyze changes over time, while those attempting to compare groups used bar graphs or pie charts. Each student explored and utilized various visualization formats based on their analytical goals, developing their own unique data representation methods.

Drawing Conclusions and Presenting

Students evaluated the validity of their hypotheses and drew conclusions based on the analyzed data. During this process, they were asked to comprehensively consider the collected data patterns, inter-variable relationships, and

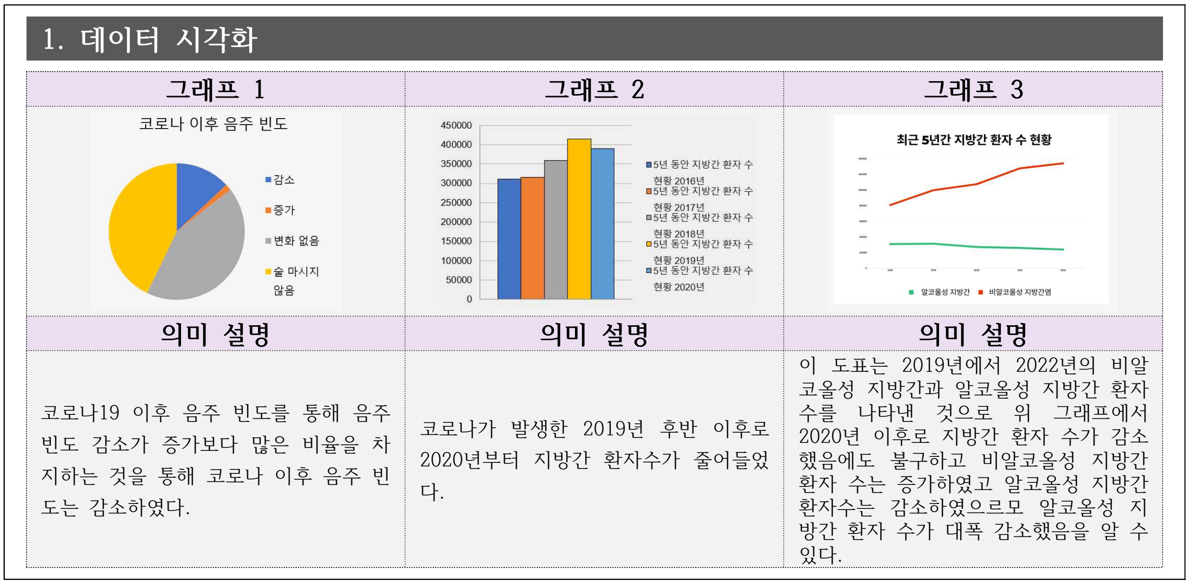

Fig. 4. Expressing data in various visualization forms

significant trends revealed in the visualization results to determine whether their hypotheses were supported or refuted. Fig. 5 shows an example of the conclusion-drawing stage.

Most students simply listed visual materials or repeatedly presented numerical results, and their implications were often limited to general ones. This is related to the classroom context, which lacked consistent practice and feedback on relevant elements throughout the inquiry process. In other words, these higher-order thinking-based constructive activities are difficult to internalize through simple activity guidance alone. This suggests that a process of gradually learning the meaning and constructive principles of each element from the beginning of the class, followed by repeated training through formative assessment and feedback, is necessary.

1. 가설 검증 및 탐구 학습 주제의 결론 도출하기

a. 데이터의 변화 양상
- 수두 발병수는경기(7,505건), 서울(3,135건) 등 인구 밀집 지역에서 높았으며, 세종(217건), 대전(671건) 등에서는 상대적으로 낮았다.
- GRDP증감률은 충북(-0.4), 전북(-0.2) 등 일부 지역에서 낮은 성장률을 보였으며, 이 지역들에서는예방접종률이평균보다 낮거나, 발병률이 높은 경향이 나타났다.

b.상관관계 및 인과관계 판단 근거
- GRDP 성장률이 낮은 지역에서 예방접종률이다소 낮거나 발병률이 높은 경향이 있는 것으로 보아, 사회경제적 수준과 감염병 대응 수준 간에는 약한 상관관계가 존재함을 추론할 수 있다.
- 다만, 인구 밀집도, 보건 인프라, 교육 수준 등 다른 변인들도 영향을 줄 수 있어, 인과관계를 단정짓기엔 한계가 있다.

c.가설의 타당성 판단
- 가설은 일정 부분 타당성이 있으며, 사회경제적 수준이 낮은 지역에서 예방접종률이낮거나 감염병 발병이 높은 경향이 확인되었다.
- 그러나 예외 지역도 있고, 인과관계를 증명할 만큼 결정적인 데이터는 부족하므로, 부분적으로 지지할 수 있는 가설로 판단된다.

2. 앞으로의 해결 과제나 시사점 제시하기

a. 해결되지 않은 문제 / 추가 검토 사항
- GRDP 외에도 인구 수, 인구밀도, 의료 접근성, 교육 수준 등 다양한 변수들을 종합적으로 분석할 필요가 있다.
- 수두 이외의 감염병(예: 인플루엔자, 홍역 등)에 대해서도 같은 분석을 적용해보면 일관된 경향성이 나타나는지 확인이 필요하다.

b.이 결론의 영향력
- 공공보건 정책 수립 시, 단순한 예방접종 권장만으로는 충분하지 않으며, 지역 간 사회경제적 격차 해소와 병행되어야 함을 시사한다.
- 의료 소외 계층에 대한 맞춤형 보건 교육 및 접종 지원의 필요성을 제기한다.

c.실천 방안
- 저소득 지역이나 의료 사각지대에 대한 예방접종 지원 확대, 지역 보건소와 연계한 교육 프로그램 강화가 필요하다.
- 학생이나 시민 대상 지역 간 건강 불균형 교육 프로그램을 실시해 건강 형평성에 대한 인식을 높일 수 있다.

d.시사점 관련 근거
- 질병관리청의 감염병 통계에 따르면, 예방접종률이높을수록 해당 질병의 유병률이 낮은 경향이 나타난다.
- 한국보건사회연구원은 사회경제적 배경이 건강 형평성에 중대한 영향을 미친다고 분석함.

Fig. 5. Student examples of drawing conclusions and presenting implications

Some students presented well-supported arguments based on the data and drew conclusions and implications using the clear framework provided by the teacher (Figure 5). Rather than merely listing numerical results, they identified meaningful changes or trends in the data and linked these patterns to their research questions, producing persuasive interpretations. In other words, beginning with core data, examining causal relationships, and gradually constructing conclusions and implications—while also considering limitations and possible alternatives—illustrate how conclusion-drawing skills and higher-order thinking can be effectively fostered through structured learning and iterative feedback.

Data Literacy and Class Satisfaction

To verify the effectiveness of an exploratory course program for big data analysis, we examined changes in data literacy. Table 3 presents the results of a t-test on data literacy.

As shown in Table 3, post-test scores significantly increased compared to pre-test scores in all subcategories ($p<0.05$). In the statistical analysis domain, improvements were observed in "Data Understanding," "Data Collection and Preparation," and "Data Analysis." Furthermore, in the data utilization and communication domain, statistically significant improvements were observed in "Data Presentation," "Data Evaluation," and "Data-Based Communication"($p<0.05$). The overall data literacy score improved from 23.29 points before the class to 25.19 points after the class, a statistically significant difference ($p<0.05$).

Table 3. Pre- and post-instruction data literacy results by subcategory

Domains	Sub-domain	Pre-test		Post-test		*t*-value
		M	SD	M	SD	
Statistical analysis	Understanding Data	3.81	0.70	4.17	0.66	6.06*
	Data Collection and Preparation	3.93	0.72	4.23	0.62	4.93*
	Data Analysis	3.71	0.81	4.17	0.67	6.27*
	Sub-total	11.45	2.23	12.57	1.95	6.69*
Data-driven communication	Data Presentation	4.10	0.67	4.31	0.61	3.28*
	Data Evaluation	3.76	0.77	4.10	0.71	5.09*
	Data-Driven Communication	3.98	0.72	4.21	0.64	3.60*
	Sub-total	11.84	2.16	12.62	1.96	5.04*
Total		23.29	4.39	25.19	3.91	6.21*

*$p<0.05$

Meanwhile, a satisfaction survey on the big data analysis class was conducted, and the results are presented in Table 4.

Table 4 shows that overall class satisfaction was high. The average total score was 64.16 out of 75. By evaluation item, the highest positive response was given to "the relevance of learning objectives and course content," followed by "achievement of learning objectives" and "clarity of course objectives."

Satisfaction with the course content, at 21.37 points, was relatively higher than satisfaction with the course objectives. Students responded that the course content was engaging and relevant to real-life situations, and they also positively evaluated the course content as "achievable by students."

Table 4. Descriptive statistics on class satisfaction

Items		M	SD	DoF	Min	Max
In struction objectives	· The course's learning objectives provided guidance on what to focus on.	4.23	0.66	124	3	5
	· The course's learning objectives were relevant and important to me.	4.17	0.70	124	2	5
	· The course's learning objectives were well-suited to my level.	4.10	0.76	124	2	5
	· The course's learning objectives and content were well-aligned.	4.33	0.64	124	3	5
	· The course achieved the learning objectives as planned.	4.28	0.66	124	3	5
Sub-total		21.11	3.42	620	13	25
In struction content	· The content learned in class was well organized.	4.34	0.70	124	2	5
	· The amount of learning in class was appropriate for me.	4.22	0.82	124	2	5
	· What I learned in class was meaningful and relevant to real life.	4.29	0.74	124	2	5
	· The learning assignments given in class were at a level that I could solve.	4.26	0.66	124	3	5
	· The content we learned in class was fun and interesting.	4.26	0.84	124	2	5
Sub-total		21.37	3.76	620	11	25
In struction effectiveness	· I believe that my ability to collect data has improved through this class.	4.31	0.71	124	2	5
	· I believe that my ability to analyze data has improved through this class.	4.36	0.68	124	3	5
	· I believe that the class has improved my ability to visualize data.	4.35	0.71	124	2	5
	· I believe that through the class, my ability to come up with solutions or implications for the topic has improved.	4.30	0.67	124	3	5
	· I think the class was worth the time and effort.	4.36	0.74	124	2	5
Sub-total		21.68	3.51	620	12	25
Total		64.16	10.69	1860	36	75

Satisfaction with the effectiveness of the course, at 21.68 points, was slightly higher than satisfaction with the course objectives and course content. Students found the course experience worthwhile and assessed that their data analysis skills had improved. They also responded positively to "improved data visualization and collection skills."

Discussion

This study examined the implementation of a big data–based science inquiry class for high school students, focusing on changes in data literacy and course satisfaction. The findings are discussed as follows.

First, students frequently struggled to formulate verifiable inquiry questions. This difficulty is likely attributable to their limited experience in generating inquiry questions independently (Kwon et al., 2008; Shin, 2007). In typical high school science classes, students primarily complete problems presented in textbooks or worksheets, leaving few opportunities to develop their own inquiry questions and refine them into analyzable problems. As a result, while students expressed interest in connecting personal interests or societal issues to inquiry topics, they often found it challenging to translate these questions into scientifically analyzable problems.

In particular, big data–based inquiry using public datasets requires that questions be formulated with consideration for data availability and analyzability (Laney, 2001; Manyika et al., 2011). Simple curiosity-based questions are therefore difficult to operationalize. When teaching inquiry question formulation, teachers should encourage students to define problems in terms of the relationship between manipulated and dependent variables, ensuring that the testability of the problem is clearly recognized (Germann & Aram, 1996).

Second, students experienced difficulties during the data collection stage, likely due to insufficient procedural thinking skills required for effective inquiry design. This aligns with previous research indicating that students with limited experience in experimental design struggle with planning experiments (Kranz et al., 2023). The open-ended nature of the data collection planning in this study further exacerbated the challenge. Future lesson designs should provide structured guidance materials that explicitly illustrate the relationships among manipulated variables, dependent variables, data sources, and data types from the outset of the lesson.

Third, during data collection, students frequently reset their collection path or relied on pre-visualized data when they could not locate the desired datasets on public data portals or statistical websites. This behavior indicates a lack of understanding regarding the educational significance of directly collecting and refining raw data. Teachers should provide detailed information about the required variables and data types in advance and emphasize the importance of data cleaning and preparation, thereby enhancing students' data collection skills.

Fourth, in the conclusion-drawing stage, students' data interpretation and logical justification were often inadequate. Although students successfully visualized the collected data, they frequently failed to fully interpret the implications. This pattern is consistent with previous findings (Yang et al., 2010; Yoo et al., 2023). Additionally, students tended to ignore data that contradicted their hypotheses, a phenomenon documented in prior research (Chinn & Brewer, 1993; Kunda, 1990), which limited their ability to draw fully evidence-based conclusions. Therefore, it is necessary to provide guidance that enables students to make evaluations based on experimental results or data.

Conclusion

This study aimed to design a science inquiry class program for high school students utilizing public COVID-19–related big data and to evaluate its educational effectiveness and practical applicability. The following conclusions can be drawn from the results:

First, the big data–based inquiry class had a positive impact on students' data literacy. Significant improvements were observed, particularly in the areas of data analysis and evaluation, demonstrating the program's effectiveness in fostering integrated data literacy skills, including the abilities to read, analyze, and communicate data.

Second, direct teacher guidance and feedback were critical throughout the inquiry process. Students encountered challenges with higher-order thinking tasks, such as formulating testable inquiry questions, planning data collection, and drawing logically justified conclusions. During the inquiry question formulation stage, students selected social issues aligned with their personal interests but struggled to translate these topics into scientifically investigable questions or refine them into testable questions involving manipulated and dependent variables.

During the data collection stage, some students relied on pre-processed visualizations rather than raw data, while others were unable to identify relevant variables, resulting in incomplete data collection. In the conclusion-drawing stage, common errors included merely summarizing analysis results or interpreting correlations as causal relationships. These findings suggest that teachers should provide structured thinking frameworks, offer specific feedback at each stage, and incorporate repeated, scaffolded practice and long-term training to support higher-order thinking skills.

Based on the results of this study, inquiry activities utilizing public big data have the potential to be expanded across diverse topics. Future research should also explore instructional methods that move beyond quantitative big data analysis to include qualitative approaches, further enriching students' data literacy and scientific inquiry skills.

References

Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108, 480-498.

Bae, H. (2019). Educational implications of data literacy in social studies. Theory and Research in Citizenship Education, 51, 95-120. http://dx.doi.org/10.35557/trce.51.1.201903.004

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. https://doi.org/10.3102/00346543063001001

D'Ignazio, C., & Klein, L. F. (2020). Data Feminism. MIT Press.

Germann, P. J., & Aram, R. J. (1996). Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. Journal of Research in Science Teaching, 33(7), 773-798.

Hwang, D. (2018). Study of Analysis on the Influence of Unplugged Computing Education on the Learning Interest and Recognition of Satisfaction: Focusing on North Korean Refugee Adolescents in a South Korean Alternative School (Unpublished Master's Thesis). Kongju National University, Kongju, South Korea.

Kim, M. (2019). The High School Students' Attitude Toward the Flipped Learning in English Class and its Effect on Academic Achievement Via a Mediating Effect of the Student Satisfaction (Unpubished Doctoral Dissertation). Seoul National University, Seoul, South Korea.

Kranz et al., 2023) Kranz, J., Baur, A., & Moeller, A. (2023). Learners' challenges in understanding and performing experiments: A systematic review of the literature. Studies in Science Education, 59, 321-367. https://doi.org/10.1080/03057267.2022.2138151

Kusuma, S., & Viswanath, D. K. (2018). IOT and big data analytics in e-Learning: A technological perspective and review. International Journal of Engineering and Technology, 7, 164-167.

Kwon, Y., Lee, J., Lee, I., Kim, Y.(2008). Development of the quotient equation of questioning ability in the analysis of secondary pre-service teachers' scientific questions on the biological phenomena. Secondary Education Research, 56, 553-579. http://dx.doi.org/10.25152/ser.2008.56.2.553

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note.

Mandinach, E. B., & Gummer, E. S. (2016). Data Literacy for Educators: Making it Count in Teacher Preparation and Practice. Teachers College Press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

Ministry of Education (2022). Science Curriculum. The Notification of MOE 2022-74: The Separate Volume No. 9, Ministry of Education.

OECD (2018). The Future of Education and Skills: Education 2030. OECD Publishing.

Popchev, I., Orozova, D., & Stoyanov, S. (2019). IoT and big data analytics in E-Learning. In 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE) (pp. 1-5). IEEE. DOI: 10.1109/BdKCSE48644.2019.9010666

Shin, D. (2007). The relationships between the patterns of elementary school teachers` explanations and the patterns of elementary school students` questions on scientific phenomena. Journal of Korean Elementary Science Education, 26, 149-160.

Shin, D. H., & Choi, H. J. (2021). Developing a big data literacy program for gifted science high school students. Journal of Science Education, 45, 123–140.

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. http://dx.doi.org/10.14697/jkase.2020.40.6.657

Song, Y., Song, D., Kim, Y., & Lim, C. (2021). A developmental study of an instructional model and strategies for data-driven debate (DDD) to improve data literacy. Journal of Educational Technology, 37, 943-982. http://dx.doi.org/10.17232/KSET.37.4.943

Swan, K., Vahey, P. J., Rafanan, K., & Stanford, T. (2009). Challenges to cross-disciplinary curricula: Data literacy and divergent disciplinary perspectives. In Annual Conference of the American Educational Research Association. https://www.sri.com/sites/default/files/publications/aera_2009_ twd.pdf

Velarde, G. (2020). Artificial intelligence and its impact on the fourth industrial revolution: a review. International Journal of Artificial Intelligence & Applications (IJAIA), 10, 41-48. https://doi.org/10.48550/arXiv.2011.03044

Yang, I., Lim, S. & Yun, S. (2010). The evidence evaluation process of gifted elementary school students in science laboratory activities. The Journal of the Korean Society for the Gifted and Talented, 9, 27-49.

Yoo, D., Lim, S., & Kim, Y.(2023). Analyses of question types and evidence evaluation types in big data analysis for 7th grade students. Brain, Digital & Learning, 13, 383-395. https://doi.org/10.31216/BDL.20230024

Authors' Information

Bo-kyung Kim: Hamji High School, Teacher, First Author. https://orcid.org/0009-0000-4641-0636

Youngshin Kim: Kyungpook National University, Professor, Corresponding Author. https://orcid.org/0000-0001-5938-5679