International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 6
ISSN: 2088-5334

Development of Data Literacy Competency System for K-12

JaeRi Jeong [a], Youngjun Lee [a,*]

[a] *Department of Computer Education, Korea National University of Education, Cheongju-si, Chungcheongbuk-do, Republic of Korea*
Corresponding author: [*] yjlee@knue.ac.kr

***Abstract*—This study focuses on developing a comprehensive data literacy competency system for K-12 education in South Korea, addressing the critical need for integrated data literacy in an increasingly data-driven society. Recognizing data literacy as a multifaceted domain essential across various disciplines, we explore its nuanced sub-competencies tailored for students at different educational levels. Our research methodology involves a qualitative content analysis of the national curriculum, supplemented by a Delphi survey with educational experts to refine and validate the identified competencies. We present a structured competency system, categorizing data literacy into four main areas: Foundation of Data, Data Collection and Management, Data Analysis and Techniques, and Data Results Utilization. Each area encompasses specific competencies and detailed competency statements, differentiated by educational level to align with students' developmental stages. The study's outcomes offer a systematic approach to embedding data literacy in the curriculum, emphasizing the importance of these skills in the digital age. Our proposed system is aligned with the national curriculum and considers the broader international context, making it relevant for global K-12 educational settings. Additionally, the system addresses the diverse needs of students, ensuring that all learners acquire essential data literacy skills. This research contributes to the field by providing a concrete framework for data literacy education, highlighting its significance in nurturing informed, capable citizens in a data-centric world. We aim to effectively prepare students to navigate and contribute to an increasingly complex digital landscape by fostering these competencies.**

***Keywords*— Data literacy; education in Korea; competency.**

Manuscript received 15 Oct. 2023; revised 14 Feb. 2024; accepted 20 Nov. 2024. Date of publication 31 Dec. 2024.

I. INTRODUCTION

In contemporary discourse, data has garnered recognition as the crude oil and primary capital of the Fourth Industrial Revolution [1]. This attribution arises from the foundational role of data in developing all facets of information and communication technologies. The ubiquitous data generation in our surroundings manifests in its accumulation across diverse contexts, serving as the cornerstone for value creation in myriad fields [2]. In the study, an epoch that unfolds in a data-driven society has been presented where the flux of operations is inexorably tied to data.

Navigating and prospering within this data-centric milieu necessitates a suite of competencies encompassing comprehension, utilization, and evaluation of the data resource. Frank and Walker [3] underscored the pivotal role of data literacy in societal membership, positing that individual's adept at creating, controlling, and understanding data accrue power, while those deficient in such literacy risk marginalization from societal participation. Data literacy embodies the capacity to harness data for individual purposes [4], encompassing technical proficiencies in data handling and analysis and the ability to objectively interpret and critically evaluate social phenomena informed by data-driven insights [5].

Recently, the term 'literacy' has been expanded by adding a specific field in front of it, referring to 'basic knowledge related to that field'[6]. Consequently, literacy has started to appear in various forms. Examples include information literacy, digital literacy, and AI literacy. Among these, data literacy has emerged as a critical area, underscoring the need for policies and education that enable the general public to acquire basic knowledge about data. Consequently, a pressing imperative emerges for formulating policies and educational initiatives to facilitate the accessible and efficacious utilization of data capital by ordinary citizens for their pursuits [7].

Data literacy permeates various disciplines, reflecting its utility across diverse fields of inquiry. Research on data literacy education spans multiple domains, including library and information sciences, mathematics, business studies, computer science, and humanities and social sciences [8].

Distinct interpretations and emphases on data literacy emerge within each discipline, underscoring its multifaceted nature [9]. Consequently, data literacy can be conceptualized as a comprehensive domain that branches into sub-disciplines contingent upon contextual considerations [21]. Therefore, a nuanced exploration of the sub-competencies of data literacy becomes imperative [10].

Wolff et al. [11] delineated a data-literate citizen as one capable of making informed decisions grounded in data within a data-driven society or critically evaluating and discerning data analysis outcomes. They categorized these citizens into four archetypes, as illustrated in Fig. 1. The 'Reader' type embodies fundamental skills essential for the general populace. In contrast, the 'Scientist' type epitomizes specialized knowledge and expertise. Individuals occupying each type exhibit disparities in knowledge breadth and focus.

Nevertheless, 'Readers' possess foundational competencies enabling them to ascend to higher tiers of proficiency as per individual exigencies. This adaptability stems from their capacity to comprehend, interpret, and evaluate data, thus laying the groundwork for acquiring advanced knowledge and complex skills [11]. This underscores the necessity of disaggregating data literacy into granular sub-competencies and hierarchical levels, facilitating a more nuanced understanding and targeted development of competencies.

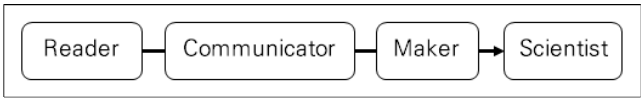

Fig. 1 The Type of Data Literate Citizen [11]

To date, we have explored the imperative of data literacy education and the necessity of segmenting it into more nuanced sub-competencies and levels, catering to diverse individuals within specific contexts. Globally, the trend in data literacy education leans towards initiating it from K-12 levels [12]. This approach is paramount, particularly in the era of AI, where data literacy becomes an essential skill. In South Korea, the K-12 national curriculum underscores the significance of data literacy, acknowledging its importance across the entire educational spectrum.

However, it is crucial to note that while information subjects, including data literacy, are part of the K-12 curriculum, their mandatory implementation is confined to middle schools [13]. This structure results in elementary and high school students receiving a different formal education in data literacy than their middle school counterparts. This gap in the curriculum highlights a significant challenge: achieving a comprehensive, nationwide standard in data literacy education is needed by the limited mandatory inclusion of information education at certain school levels.

A potential solution to bridge this gap is the integration of data literacy across various subjects, not just within the confines of designated information classes. However, this broadened approach necessitates a unified and precise definition of data literacy. A consistent understanding across disciplines is essential for effectively implementing data literacy education [11]. This inclusive approach ensures that data literacy is not merely a specialty of a few but an integral component of a holistic educational framework, equipping students to navigate and contribute in a data-centric world proficiently.

In this study, our primary aim is to develop a comprehensive data literacy education competency system tailored to the South Korean K-12 education curriculum. To achieve this, we have delineated the following research questions:

a. What constitutes the perspective of data literacy for K-12 students? This question seeks to identify the foundational aspects of data literacy that are relevant and appropriate for K-12 students.
b. What specific data literacy competencies should be imparted to K-12 students? This involves delineating the competencies necessary for students at different educational stages and ensuring that the curriculum aligns with their developmental needs and capacities.
c. What challenges and objectives are necessary to achieve these competencies at each school level? This question aims to map out the distinct objectives and challenges at different educational levels, enabling a targeted approach to curriculum development.

To address these questions, the study will examine the characteristics of data literacy concepts commonly emerging in K12 data literacy education research. Additionally, through a qualitative content analysis of the national curriculum, we aim to extract and define the specific competencies necessary for effective data literacy education in South Korea.

II. Materials and Method

A. Research Procedure

The research procedure for developing a data literacy competency system, as depicted in Table I, involved two primary stages: the extraction of specific data literacy competencies suitable for K-12 students and the development of school-level competency statements based on these competencies.

TABLE I
RESEARCH PROCEDURE

Stage	Process
Extraction of competencies	Analyzing the curriculum Delphi survey with experts
Development of competency statements	Analyzing prior research Drafting statements Refining through Delphi surveys Compile the competency system

Initially, to identify data literacy competencies that align with K-12 students' levels, we analyzed content from the information education curriculum most relevant to data literacy competencies. The preliminary competencies extracted from this analysis were then subjected to a Delphi survey among expert groups to ensure validity and finalize these data literacy competencies.

Subsequently, the development of school-level competency statements commenced. This involved analyzing prior research to determine content elements and proficiency levels and creating an initial draft. The draft of these competency statements underwent two rounds of Delphi survey with expert groups for refinement. The finalized competency system and definitions of the established

competencies were then systematically compiled and presented [14].

B. Deriving Detailed Competencies of Data Literacy

To extract specific data literacy competencies necessary for K-12 students in Korea, a content analysis of the 2022 revised Korean educational curriculum was conducted. The focus was on the content included in the Information subject of this curriculum.

While not existing independently, the 'Digital Society and Artificial Intelligence (AI)' unit in the Practical Arts subject for elementary schools is integrally linked to the middle school Information curriculum. Therefore, the study also encompassed the 'Digital Society and AI' unit and 'Information Education' supplemental materials designed for autonomous hours [15]. For middle school, the analysis included data literacy-related content in the 'Information' subject [16], and the high school curriculum encompassed subjects like 'Information,' 'Foundations of Artificial Intelligence,' 'Data Science,' and 'Software and Life'[17].

The content analysis method employed was the deductive qualitative content analysis approach, as Elo and Kyngäs [18] suggested. This approach involves developing an analysis framework based on existing theory or using a well-structured existing framework for coding the content. The data literacy framework developed by Ridsdale et al. [19] was utilized in this study. They defined data literacy through five components (data, collection, management, evaluation, application) and identified 23 sub-competencies within these, along with 64 examples of the expected knowledge and tasks associated with each competency (Table II). This framework, widely cited in numerous studies [20]–[22], was derived from a broad qualitative analysis of formal and informal data literacy literature.

However, the tasks and knowledge outlined in this framework make it somewhat ambiguous to identify each competency's specific scope and elements. The tasks are predominantly geared towards general education at the university level, targeting college graduates. This presents a discrepancy with the context of data literacy as intended in this study. Therefore, modifications were made to establish more precise criteria for the somewhat vaguely presented competencies in the framework and to revise some of the content during the coding process deemed unsuitable for our study's context [23].

TABLE II
DATA LITERACY FRAMEWORK [19]

Area	Competency
Data	- Introduction to Data
Collection	- Data Discovery and Collection - Evaluating and Ensuring Quality of Data and Sources
Management	- Data Organization - Data Manipulation - Data Conversion - Metadata Creation and Use - Data Curation, Security, and Re- Use - Data Preservation
Evaluation	- Data Tools - Basic Data Analysis - Data Interpretation - Identifying Problems Using Data
Application	- Data Visualization - Presenting Data - Data Driven Decision Making - Critical Thinking - Data Culture - Data Ethics - Data Citation - Data Sharing - Evaluation Decisions Based on Data

The coding process was conducted in three stages, as detailed in Table III. The first stage focused on extracting all relevant content without omissions. The unit of analysis was set to a single sentence, with multiple codlings applied if a sentence contained pertinent several pieces of information [24]. Details such as grade level, subject, unit, and page number were recorded along with the extracted content. Any content not categorized into competencies was marked and annotated during this stage for further review [23]. In the second coding stage, the content was filtered by categorized competencies. Commonalities were then subdivided into sub-elements. This categorization allowed for the recoding of content from the first stage that had not been classified, leading to the merging of some competencies. The third stage involved grouping the extracted competencies into higher-level areas.

TABLE III
CODING PROCESS

Stage	Method
1	Extracting all relevant content
2	Subdividing into sub-element
3	Grouping into higher-level areas

Following completing these three coding stages, a Delphi survey was conducted to establish the content validity of the identified essential competencies. The survey targeted 15 teachers specializing in computer education with training experience in AI and data education. Responses were collected on a Likert 5-point scale, and additional feedback was sought for responses indicating a score of 2 (Invalid) or 1 (Very Invalid). The Delphi survey results included calculating average values, standard deviations, Content Validity Ratio (CVR) values, and the extent of concordance and convergence in the experts' opinions [25].

C. Development of Data Literacy Competency Statements

Developing competency statements is crucial in structuring the data literacy competency system. These statements are essential components of competency-based curricula, as outlined in the ACM's Computing Curricula 2020: Paradigms for Global Computing Education (CC2020). Competency statements provide examples of expected performance outcomes for students, succinctly describing achievable abilities and objectives and facilitating communication between teachers and stakeholders. CC2020 provides a guide for developing competency statements with the following criteria [26]:

1) Teacher's Perspective: The statements should be written from the perspective of educators conducting the training.

2) Inclusion of All Sub-Competencies: The statements must clearly articulate each sub-competency.

3) Action-Oriented Verbs: Statements should be as specific and straightforward as possible, using action-oriented verbs. Beyond mere knowledge and skills, they should explicitly articulate the particular contexts and expected outcomes. Krathwohl [27] revised educational taxonomy in the cognitive domain to describe outcome behaviors as action oriented.

4) Appropriate Student Level: The statements are tailored to what students at each educational level can achieve. The achievement standards of the Korean national curriculum were consulted to ensure level-appropriateness and coherence. And foreign curriculum was also referred to.

This study developed competency statements for each sub-competency for three levels, corresponding to elementary, middle, and high school students [14]. To set criteria for these levels, the characteristics of proficiency levels are outlined in Dig Comp 2.1: The Digital Competence Framework for Citizens [28]. Krathwohl [27] revised educational taxonomy was also utilized for action-oriented periodical descriptions of result behaviors. The characteristics of each level, as established, are summarized in Table IIV. Level 1 focuses on understanding basic concepts and applying them in simple tasks, described as 'Remember' and 'Understand.' Level 2 involves structured, familiar tasks that allow the application of various skills. Level 3 pertains to unstructured, real-world problems where students independently choose data or methodologies and discover or create meaningful information [28].

A two-round expert Delphi survey was conducted to ensure the validity of the developed draft competency statements. The panel comprised 11 experts, including computer education teachers with experience conducting AI and data science education or training for students and teachers, and one university professor specializing in the field. The experts were briefed on the objectives, background knowledge, and the development process of the study's outputs. They were given a table detailing the specific data literacy competencies and their descriptions.

The review was conducted for each statement, with the criteria for assessment outlined in Table V. Responses for each item were collected on a 5-point Likert scale, and open-ended feedback was gathered regardless of the score. From the first round of the survey, average values, standard deviations, Content Validity Ratio (CVR), degree of opinion convergence, agreement, and stability were calculated. The CVR was determined by the proportion of panelists who agreed with a score of 4 (Valid) or 5 (Very Valid).

TABLE IV
THE CHARACTERISTICS OF EACH LEVEL

Level	Tasks	Cognitive
Level 1	Foundational concepts	Remember
	Simple task	Understand
Level 2	Well-defined and routine tasks	Apply
		Analyze
Level 3	Complex problems in the real world	Evaluate
		Create

The second Delphi survey involved the same panel of experts and focused on the revised statements reflecting feedback from the first round. For this survey, experts were provided with a summary of the first-round results, feedback, and modifications. They received a questionnaire indicating the previous average score (M) and their previous response (x) for each statement, allowing them to provide a new response (X) and additional comments. Table VII showcase a part of the questionnaire for Domain 1 statements.

TABLE V
THE FIRST DELPHI SURVEY

Criteria	1	2	3	4	5
(1) Inclusion of sub-competencies					
(2) Clarity and specificity of statements					
(3) Appropriateness of the statements for the respective educational levels					

TABLE VI
THE SECOND DELPHI SURVEY

Statements	Criteria 1			Criteria 2			Criteria 3		
	M	x	X	M	x	X	M	x	X
1-1 (Primary)									
1-2 (Middle)									
1-3 (High)									

M: the previous average score
x: previous response, X: new score

III. RESULTS AND DISCUSSION

A. Sub-Competencies of Data Literacy

To identify necessary data literacy capabilities for K-12 students, a qualitative content analysis of Korea's 2022 revised Information Education Curriculum was conducted. This analysis utilized Ridsdale et al.'s [19] data literacy framework as the deductive framework. The coding process consisted of three stages. During the first stage, competencies that did not yield related content were considered beyond the scope of secondary education and thus eliminated from the framework. This resulted in removing competencies such as Data Citation, Data Sharing, Metadata Creation and Use, Data Curation, Security, and Re-Use, which were classified as advanced in the original framework. Critical Thinking competency, while essential for data literacy, transcends sub-competencies and is fundamental across all literacy skills [22]. Recognizing it might represent a higher-order skill within data literacy, it was deemed an independent competency and thus excluded.

The second stage involved categorizing content into sub-elements based on competencies. Common themes in these sub-elements allowed for the recoding of initially unclear content. It led to the merging or modifying of some competencies (see Fig. 2).

For instance, the 'Evaluating and Ensuring Quality of Data and Sources' competency was merged with the 'Data Discovery and Collection' competency, as it represented a skill that could be encompassed within it. Similarly, the 'Identifying Problems Using Data' competency was merged into the 'Data Interpretation' competency capability.

The 'Data Organization' and 'Data Manipulation' competencies, which were treated in the national curriculum as methods for data management, were merged into the 'Data Management' competency. The 'Data Conversion' and 'Data

Preservation' competencies, while included in the curriculum, were integrated into the 'Data Management' competency as they were treated more as skills for efficient and secure data handling rather than in-depth technical skills.

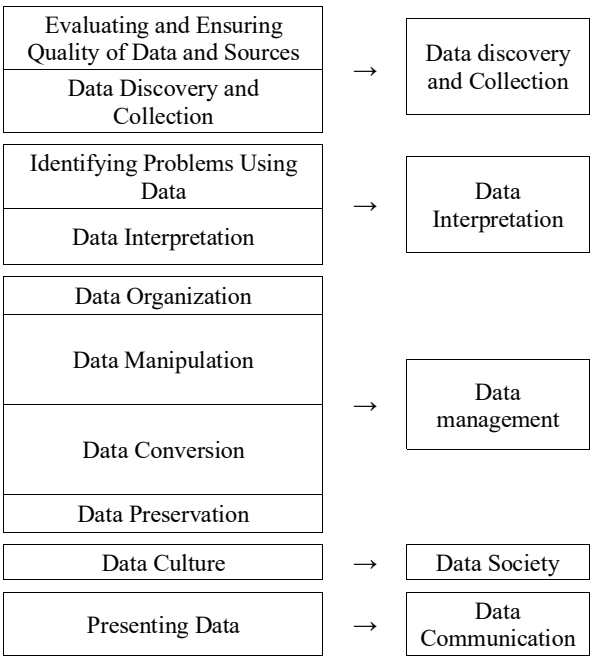

Fig. 2 Merging or Modification of Some Competencies

Additionally, some competencies were renamed to align with terms used in the Korean curriculum. 'Data Culture' competency, conceptualized as recognizing the importance of data in learning, research, and decision-making, was more commonly linked to society in Korean curricula and thus renamed to 'Data Society' competency. Similarly, the 'Presenting Data' competency, involving data for persuasive visual communication, aligned with the curriculum's emphasis on communication skills and was revised to the 'Data Communication' competency.

The derived competencies were grouped into four main areas in the third stage. The conceptual knowledge about data, data society, and data ethics competencies were categorized as the 'Foundation of data.' The data exploration, collection, and organizing competencies were grouped into 'Data Collection and Management.' Basic data analysis skills, visualization, interpretation, and analysis tools and techniques were grouped under 'Data Analysis and Technology.' Finally, competencies involving decision-making and communication using data analysis results were grouped as 'Data Result Utilization'. Consequently, the sub-competencies of data literacy were categorized into four areas and 12 sub-competencies, as outlined in Table VIIII.

A Delphi survey with 15 experts validated these competencies, as detailed in Table VIIIIII. The table presents the mean value(M), standard deviation (SD), coefficient of variation (CV), content validity ratio (CVR), concordance (CC), and convergence (Conv.) for each sub-competency. The CVR threshold value from the 15-member panel was .49, confirming the validity of all sub-competencies. However, there were suggestions to reconsider the 'Data Tools and Technique' competency as it spans various areas requiring technical skills, not limited to data analysis. It was eventually integrated into relevant competencies rather than being treated as independent. As a result, the final categorization of data literacy included four areas and 11 sub-competencies, detailed in Table IIX.

TABLE VII
DATA LITERACY SUB-COMPETENCIES DRAFT

Area	Competency
Foundation of data	Data Concepts
	Data Society
	Data Ethics
Data Collection and Management	Data Discovery and Collection
	Data Management
	Data Processing
Data Analysis and Techniques	Basic Data Analysis
	Data Visualization
	Data Interpretation
	Data Tools and Techniques
Data Results Utilization	Data Driven Decisions Making
	Data Communication

TABLE VIII
EXPERT DELPHI RESULTS FOR DATA LITERACY SUB-COMPETENCIES

Sub-competencies	M	SD	CV	CVR	CC	Conv.
Data Concepts	4.87	.35	.07	1.00	1.00	.00
Data Society	4.67	.49	.10	1.00	.80	.50
Data Ethics	4.93	.26	.05	1.00	1.00	.00
Data Discovery and Collection	4.93	.26	.05	1.00	1.00	.00
Data Management	4.80	.56	.12	.87	1.00	.00
Data Processing	4.87	.52	.11	.87	1.00	.00
Basic Data Analysis	4.73	.46	.10	1.00	.90	.25
Data Visualization	4.87	.35	.07	1.00	1.00	.00
Data Interpretation	4.67	.90	.19	.73	1.00	.00
Data Driven Decision Making	4.80	.41	.09	1.00	1.00	.00
Data Communication	4.53	.64	.14	.87	.80	.50

TABLE IX
DATA LITERACY SUB-COMPETENCY

Area	Comp.	Description
1. Foundation of Data	1-1. Data Concepts	Recognizing the necessity of digital data based on an understanding of data utilized in computing systems, various types of data, such as numerical, textual, auditory, and visual data, can be represented as digital data.
	1-2. Data Society	To recognize societal issues related to data and to comprehend fields and technologies associated with data."
	1-3. Data Ethics	Recognizing ethical issues related to data and identifying ethical concerns in the data analysis."
2. Data Collection and Management	2-1. Data Discovery and Collection	Selecting and exploring appropriate data sources aligned with the objectives and selecting and gathering necessary data. Selecting appropriate data exploration and collection tools and methods as case.
	2-2. Data Management	Creating a data table for systematic data management, organizing and storing data through manipulations such as classification and sorting as needed."
	2-3. Data Processing	Preprocessing, such as cleansing, integration, and normalization, facilitates data analysis and ensures data integrity by preventing distortion or errors.
3. Data Analysis and Techniques	3-1. Basic Data Analysis	Understanding the characteristics of data and utilizing appropriate data analysis tools and methods to generate valuable insights
	3-2. Data Visualization	Selecting and generating appropriate visual representations to understand relationships within data or discover helpful information, depending on the objectives.
	3-3. Data Interpretation	Interpreting visualized graphs to identify critical points or significant patterns.
4. Data Results Utilization	4-1. Data-Driven Decision Making	Evaluating the value of data analysis results and making rational decisions for problem-solving.
	4-2. Data Communication	Effectively utilizing data analysis results to communicate and advocate for one's opinions.

B. Development of Competency Statements

Competency statements concisely describe the abilities and objectives that can be achieved through the learning process [26]. They provide consistent information about competencies to educators implementing the curriculum. To select the content, we analyzed international data literacy educational materials and the domestic national curriculum of Korea. In this section, we introduce the detailed competencies of data literacy, explore the Korean educational curriculum to draft a preliminary version, present the results and amendments from the Delphi survey conducted on the draft, and introduce a systematically structured system with revised competency statements.

1) Analysis Curriculum for Draft: The 'Foundation of Data' area includes competencies like Data Concepts, Data Society, and Data Ethics. These competencies cover a basic conceptual understanding of data and awareness of social and ethical issues related to data. 'Data Concepts' competency involves exploring different types of data used in computing systems and understanding the digital data concepts that represent these in computing systems, recognizing the necessity of digital data [15], [29]. It also includes understanding how various data types can be used in computing systems and the methods needed.

'Data Society' competency is about recognizing social issues related to data and understanding the impact and value of data. This includes exploring examples of data-based problem-solving and understanding how data is widely used and generates value in various fields [30], [31]. 'Data Ethics' competency deals with ethical issues related to data, such as understanding the potential problems arising from data use, like bias and privacy issues, and contemplating ways to prevent or solve these issues.

'Data Collection and Management' encompasses competencies like Data Discovery and Collection, Data Management, and Data Processing. 'Data Discovery and Collection' competency is about exploring and collecting the right data for problem-solving through various methods. This competency evolves from simply collecting relevant data to being able to explore and select the right data actively [32]. 'Data Management' competency involves understanding and applying the knowledge and skills necessary to manage data systematically, from simple to large-scale datasets.

'Data Processing' competency covers the skills to manipulate data appropriately as needed, such as for data analysis tasks. This includes identifying necessary and unnecessary data in a dataset and detecting potential errors before analysis to prevent distorted results [33].

In the 'Data Analysis and Technique' area, competencies like Basic Data Analysis, Data Visualization, and Data Interpretation are included. 'Basic Data Analysis' is the ability to understand and use basic statistical skills necessary for analyzing data, along with knowledge of tools and methods for data analysis. The scope of these skills and tools can vary depending on the level. Using computing tools, primary schools focus on basic aggregate operations like sums and averages. Middle schools teach how to compute representative values like median and mode and how to select appropriate data analysis tools and methods. High schools cover topics like big data analysis and methods for analyzing unstructured data.

'Data Visualization' involves the skill of representing data in visual graph forms for easier understanding. It's essential to be able to choose the most effective graphs for representing data and understand the information that different graphs can reveal [29]. This skill helps uncover valuable insights, such as relationships between variables in the data [34]. 'Data Interpretation' is not just about reading the quantitative and

temporal ranges in visualized graphs but also recognizing patterns, rules, and trends that emerge from the overall graph.

This competency evolves from merely reading superficial information in graphs to inferring hidden patterns and relationships and discovering new problems within the data [35]–[38]. The progression of graph visualization skills is evident in the mathematics curriculum, starting with univariate graphs like bar and line graphs in elementary school, moving to bivariate graphs like scatter plots in middle school, and then to understanding features of data and identifying key variables using box plots and diagrams in high school.

The 'Data Result Utilization' area comprises Data-Driven Decision Making and Data Communication competencies. These involve using inferred information from data to choose new phenomena or problem-solving methods rationally or to present one's argument logically. The Korean language arts curriculum reflects this progression of skills, emphasizing data-based argumentation and presentation across educational levels. Elementary students learn to select data considering the audience and medium, while middle school students learn to reorganize data for more effective delivery and systematically organize content. Referencing these standards, elementary education focuses on choosing appropriate information based on the purpose and logically delivering one's opinion. In contrast, middle school education includes the ability to provide clear and consistent information using data [39], [40].

2) Delphi Survey Results: The results of the expert Delphi survey on the drafted competency statements are shown in Table X. Responses were collected on a 5-point Likert scale for three criteria: (1) inclusion of sub-competencies, (2) clarity and specificity of statements, and (3) appropriateness of the statements for the respective educational levels. Statements a, b, and c correspond to elementary, middle, and high school levels. Most statements exceeded the CVR threshold of .59 for the panel of 11 experts, but some fell short of the threshold or were suggested for further review.

The revisions based on the review comments are organized as follows: First, regarding the inclusion of specific competencies, there was feedback suggesting that the understanding of the importance and verification methods of data sources, initially included only at the high school level in the 'Data Discovery and Collection' competency, should be introduced at primary and middle school levels as well(2-1a, 2-2b in Table XI).

TABLE XII
RESULTS OF THE FIRST ROUND DELPHI SURVEY ON DRAFT

Area	Competency	Statement	Criteria (1)			Criteria (2)			Criteria (3)		
			M	SD	CVR	M	SD	CVR	M	SD	CVR
1.	1-1	1-1-a	4.64	.50	1.00	4.45	.82	.64	4.82	.40	1.00
Foundation of Data	Data Concepts	1-1-b	4.55	.69	.82	4.73	.65	.82	4.73	.65	.82
		1-1-c	4.64	.67	.82	4.73	.65	.82	4.45	.69	.82
	1-2	1-2-a	4.73	.65	.82	4.36	.92	.45	4.36	.67	.82
	Data Society	1-2-b	4.55	.69	.82	4.64	.67	.82	4.73	.65	.82
		1-2-c	4.73	.65	.82	4.73	.65	.82	4.45	.82	.64
	1-3	1-3-a	4.64	.67	.82	4.64	.67	.82	4.55	.69	.82
	Data Ethics	1-3-b	4.64	.67	.82	4.82	.60	.82	4.73	.65	.82
		1-3-c	4.82	.60	.82	4.64	.81	.64	4.45	.69	.82
2.	2-1.	2-1-a	4.64	.81	.64	4.73	.47	1.00	4.73	.47	1.00
Data Collection and	Data Discovery and	2-1-b	4.64	.81	.64	4.91	.30	1.00	4.73	.47	1.00
Management	Collection	2-1-c	4.91	.30	1.00	4.73	.65	.82	4.64	.67	.82
	2-2.	2-2-a	4.91	.30	1.00	4.82	.40	1.00	4.64	.50	1.00
	Data Management	2-2-b	4.91	.30	1.00	4.82	.60	.82	4.82	.40	1.00
		2-2-c	4.91	.30	1.00	4.82	.40	1.00	4.73	.47	1.00
	2-3.	2-3-a	4.82	.40	1.00	4.91	.30	1.00	4.55	.82	.64
	Data Processing	2-3-b	4.64	.67	.82	4.82	.60	.82	4.55	.69	.82
		2-3-c	4.91	.30	1.00	4.36	.92	.45	4.36	.67	.82
3.	3-1.	3-1-a	4.64	.67	.82	4.18	.87	.45	4.73	.47	1.00
Data Analysis and Technique	Basic Data Analysis	3-1-b	4.91	.30	1.00	4.64	.67	.82	4.45	.69	.82
		3-1-c	5.00	.00	1.00	5.00	.00	1.00	4.73	.47	1.00
	3-2.	3-2-a	4.82	.40	1.00	4.27	.90	.45	4.45	.69	.82
	Data Visualization	3-2-b	4.82	.40	1.00	4.91	.30	1.00	4.64	.50	1.00
		3-2-c	4.73	.65	.82	4.55	.69	.82	4.64	.50	1.00
	3-3.	3-3-a	4.73	.65	.82	4.55	.69	.82	4.64	.50	1.00
	Data Interpretation	3-3-b	4.73	.65	.82	4.55	.69	.82	4.64	.50	1.00
		3-3-c	4.73	.65	.82	4.55	.69	.82	4.64	.50	1.00
Data Result Utilization	4-1.	4-1-a	4.91	.30	1.00	4.55	.69	.82	4.55	.52	1.00
	Data Driven Decision	4-1-b	4.91	.30	1.00	4.55	.69	.82	4.73	.47	1.00
	Making	4-1-c	4.91	.30	1.00	4.91	.30	1.00	4.64	.50	1.00
	4-2.	4-2-a	4.82	.40	1.00	4.82	.40	1.00	4.36	.67	.82
	Data Communication	4-2-b	4.91	.30	1.00	4.73	.65	.82	4.64	.50	1.00
		4-2-c	4.91	.30	1.00	4.73	.47	1.00	4.64	.50	1.00

This modification involves progressively integrating the understanding of source importance and verification methods, starting from recognizing the importance and understanding the methods of source verification at the primary level,

exploring reliable sources for various collection methods at the middle school level, and evaluating the quality of collected data at the high school level [32],[33] [36]–[38].

Second, for specificity in the statements, there were requests for more detailed feedback on terms like 'various data forms' in Data Concepts (1-1-a), 'error values and methods to handle them' in Data Processing (2-3-c), 'performing data tasks' in Data Analysis (3-1-a), and 'simple visual graphs' (3-2-a). Consequently, these statements were revised to include examples in parentheses. For instance, 'Identifying error values, analyzing their causes, and handling them appropriately (2-3-c)' was modified to 'Identifying error values (missing values, outliers, etc.) and handling them with suitable methods (deletion, normalization, value replacement, etc.).

Third, regarding the appropriateness of the level, the capability related to 'understanding of data science' in Data Society (1-2-c) was removed as it was deemed too advanced and not well-connected to the elementary and middle school curriculum. In the data analysis domain, using box plots and histograms for analysis initially included at the high school level was deemed inappropriate as they are also part of the middle school curriculum. The content was not initially included in middle school due to the distribution of topics presented in the later years (higher grades), making it seemingly unsuitable to use these skills and knowledge. However, considering it is feasible for the middle school level and the connection with high school content, it was decided to add this content at a basic level in middle school as well [16], [38], [39].

3) Data Literacy Competency System: The results of the second Delphi survey on the revised statements indicated that all statements achieved a CVR value exceeding the threshold of .59 for the 11-member panel, with a convergence below .50, indicating consensus among the experts. Furthermore, the agreement level exceeded .75, demonstrating a solid consensus among the experts' opinions. The coefficient of variation calculated for all responses was below .50, indicating stable views among the experts, thus concluding the Delphi survey without further examination. Based on additional review comments, modifications for concise and clear communication were made, and the competency system was organized after an internal review [37]–[40].

The components of the competency system include domains, detailed competencies, descriptions of these competencies, and competency statements differentiated by educational levels. During this process, a structured table was developed to present the descriptions of competencies and their connections across different school levels (Table XIII). This approach led to the development of a competency system that considers the specific competencies and the educational level at which they are introduced.

TABLE XIVXV
DATA LITERACY EDUCATION COMPETENCY SYSTEM FOR K-12

Area	Competency	Description	Statement		
			Primary	Middle	High
1. Foundation of Data	1-1. Data Concepts	Recognizing the necessity of digital data based on an understanding of data utilized in computing systems, various types of data, such as numerical, textual, auditory, and visual data, can be represented as digital data.	Explore various types of data, including text, images, sound, and video, learn to differentiate between analog and digital data, and recognize the importance of digital data.	Build on an understanding of the digital representation of various data types and use computing tools to convert different types of data into digital form.	Comprehend the concept and characteristics of big data and, with an understanding of structured and unstructured data, use computing tools to convert unstructured data into structured data.
	1-2. Data Society	To recognize societal issues related to data and to comprehend fields and technologies associated with data."	Explore where and how data encountered in daily life is collected and utilized and recognize the value and importance of data.	Investigate examples of how data is used in various fields, including technical and economic aspects, and discuss its individual and social impacts from both positive and negative perspectives.	Acknowledge the influence of data across diverse fields and envision the future of society as it evolves due to the impact of data.
	1-3. Data Ethics	Recognizing ethical issues related to data and identifying ethical concerns in the data analysis."	Explore ethical issues surrounding data in their environment and recognize the importance of data protection.	Understand the impact of data on artificial intelligence learning, identify problems arising from it, and analyze their causes.	Reflect on social and ethical issues in data analysis and devise and present solutions to these problems.
2. Data Collection and Management	2-1. Data Discovery and Collection	Selecting and exploring appropriate data sources aligned with the objectives and selecting and gathering necessary data. Selecting appropriate data exploration and collection tools and methods as case.	Recognize the importance of sources for exploring and collecting data and understand how to verify these sources. Distinguish between relevant and irrelevant data to a given problem and collect only the necessary data.	Recognize various methods and sources for data collection and use the most suitable method depending on the purpose. Understand the concepts of public and private data and explore reliable data on open data platforms.	Plan data sources and collection methods based on the data characteristics needed for problem-solving. Evaluate data quality regarding accuracy, completeness, and security from reliable sources and determine and collect useful data for problem resolution.
	2-2. Data Management	Creating a data table for systematic data management, organizing and storing data through manipulations such as	Categorize collected data into variables based on criteria, organize it into tables, and save it as files. Search for necessary data within the stored files.	Filter and group data according to specific criteria to emphasize or clarify certain parts of the data.	Operate to search for or extract necessary data and integrating two datasets for efficient management.

Area	Competency	Description	Statement Primary	Middle	High
		classification and sorting as needed."			
	2-3. Data Processing	Performing preprocessing such as cleansing, integration, and normalization to facilitate data analysis and ensure data integrity by preventing distortion or errors.	Organize data for analysis, discarding unnecessary data or consolidating necessary data from two sources into one file.	Clean data into a consistent format for analysis. Understand the concepts of values that can cause errors in data analysis results(such as missing values and outliers) and explain the causes and impacts of these values.	Preprocess unstructured data for analysis using computing tools. Identify errors (such as missing values and outliers) and appropriately address them using suitable methods (like deletion, normalization, or replacement) depending on the situation.
3. Data Analysis and Techniques	3-1. Basic Data Analysis	Understanding the characteristics of data and utilizing appropriate data analysis tools and methods to generate valuable insights	Learn how to calculate the mean and understand. Perform basic operations (such as summation and average) using computing tools.	Select appropriate data analysis tools and methods based on their objectives, and compare the results of applying different analytical methods to the same data set. Use computing tools to calculate representative values of data (such as mean, median, and mode).	Understand the principles of unstructured data analysis methods and apply them in comparison with structured data analysis method.
	3-2. Data Visualization	Selecting and generating appropriate visual representations to understand relationships within data or discover useful information, depending on the objectives.	Understand simple visual graphs (such as bar, line, and pie charts) representing univariate data and differentiate between them based on their intended use.	Visualize bivariate data using scatter plots to reveal correlations between variables. Use histograms or box plots to visualize data distribution.	Use scatter plots to discern trends and tendencies within the data, gaining insights into relationships and patterns. Utilize scatter plots to represent three variables through variations in color and shape or to identify groups within the data.
	3-3. Data Interpretation	Interpreting visualized graphs to identify key points or significant patterns.	Identify and interpret data points in visual graphs, such as prominent features or the extent of value distribution.	Read basic statistical information from visual graphs about datasets and infer relationships between variables.	Comprehend data characteristics through diagrams representing distributions and identify key variables. Interpret results of regional data analysis related to population structure, environmental issues, or regional development and infer new problems.
4. Data Results Utilization	4-1. Data Driven Decision Making	Evaluating the value of data analysis results and making rational decisions for problem-solving.	Select appropriate analysis results to convey their opinions effectively using data	Choose rational problem-solving methods based on data analysis results.	Compare and analyze the potential social impacts of problem-solving methods based on data.
	4-2. Data Communication	Effectively utilizing data analysis results to communicate and advocate for one's opinions.	Describe the process of their data work and communicate their opinions based on information derived from data analysis.	Considering the level of the audience, use appropriate data visualizations and clearly and consistently present their intended message.	Reorganize the results of their data analysis work to be easily understood by the audience and present them systematically and logically.

IV. Conclusion

This research was initiated to educate K-12 students about the fundamental data literacy necessary for a data-driven society. Data literacy, applicable not only in information subjects but across most disciplines, varies in its focus and specific competencies depending on the subject [9]. Hence, it's essential to articulate a consistent and concrete set of data literacy competencies for a common understanding.

In pursuit of this goal, our research developed a competency system for data literacy necessary for students at all levels of K-12 education. We extracted detailed competencies and developed statements describing the learning process to achieve each competency, which were then organized into a comprehensive system. The proposed competency system breaks new ground by categorizing data literacy into distinct, age-appropriate segments, encompassing the Foundation of Data, Data Collection and Management, Data Analysis and Techniques, and Data Results Utilization.

The conclusions drawn from this research are as follows: First, we found that data literacy is not a static concept but rather an evolving skillset, crucial at every educational stage and adaptable to students' diverse and changing needs. By meticulously structuring competencies and competency statements, the study bridges an essential gap in the existing curriculum, providing a roadmap for educators and policymakers to foster data-savvy individuals from a young age.

Secondly, more far-reaching, by instilling data literacy competencies at an early age, we pave the way for future generations to not only navigate but also shape a world where data is integral. The competency system serves as a robust framework for developing curricula that are aligned with current educational standards and forward-looking in their approach to data literacy.

Looking ahead, our study opens avenues for further research. The practical implementation of this competency system and its effectiveness in real-world educational settings need empirical verification. Its adaptability and applicability in international contexts also present another fertile ground for exploration. As data literacy evolves with technological advancements, continuous updates and refinements to the competency system will be essential to keep pace with the changing landscape. In conclusion, this research contributes a structured framework for K-12 data literacy education. It sparks a dialogue on the need for ongoing evolution in educational approaches to meet the demands of a data-centric world.

ACKNOWLEDGMENTS

This paper is a revised and expanded version of a Master's thesis originally submitted to Korea National University of Education.

REFERENCES

[1] T. Kim, C. Kim, and H.-G. Yoon, "Exploring the Feasibility of Data Literacy as a Liberal Arts Course," *The Korean Data Analysis Society*, vol. 22, no. 6, pp. 2475–2487, Dec. 2020, doi:10.37727/jkdas.2020.22.6.2475.

[2] "Data Literacy Framework," Queensland Government, Queensland, Australia, 2021. [Online]. Available: https://education.qld.gov.au/initiativesstrategies/Documents/data-literacy-framework.pdf.

[3] M. Frank and J. Walker, "Some Key Challenges for Data Literacy," *The Journal of Community Informatics*, vol. 12, no. 3, Sep. 2016, doi:10.15353/joci.v12i3.3288.

[4] W. Choi, "Data literacy education and 'Songnamjapji'," *Korean Studies Quarterly*, vol. 44, no. 4, pp. 187–220, 2021.

[5] J.-M. Lee, "Re-approach to the Concept of Data Literacy and Its Application to Library Information Services," *Journal of the Korean Society for Library and Information Science*, vol. 53, no. 1, pp. 159–179, Feb. 2019.

[6] W. J. Lee, H. J. Baek, "A Study on the Research Trends of Political Education Through the Keyword Network Analysis," *Journal of Law-Related Education*, vol. 12, no. 3, pp. 73 − 93, Dec. 2017, doi:10.29175/klrea.12.3.201712.73.

[7] "The Data Manifesto," Royal Statistical Society (RSS), 2019. [Online]. Available: https://rss.org.uk/policy-campaigns/policy/our-data-manifesto/.

[8] H. Lee and S. Han, "An Analysis of Data Science Curriculum in Korea," *Journal of the Korean Society for Library and Information Science*, vol. 54, no. 1, pp. 365–385, Feb. 2020.

[9] K. Choi and D.-s. Cho, "Re-conceptualization of data literacy reflecting the expanded data characteristics and context," *Informatization Policy*, vol. 30, no. 3, pp. 49–68, 2023.

[10] B. Ghodoosi, T. West, Q. Li, G. Torrisi-Steele, and S. Dey, "A systematic literature review of data literacy education," *Journal of Business & Finance Librarianship*, vol. 28, no. 2, pp. 112–127, Feb. 2023, doi: 10.1080/08963568.2023.2171552.

[11] A. Wolff, D. Gooch, J. J. Cavero Montaner, U. Rashid, and G. Kortuem, "Creating an Understanding of Data Literacy for a Data-driven Society," *The Journal of Community Informatics*, vol. 12, no. 3, Aug. 2016, doi: 10.15353/joci.v12i3.3275.

[12] A. Grillenberger and R. Romeike, "Developing a theoretically founded data literacy competency model," *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, pp. 1–10, Oct. 2018, doi: 10.1145/3265757.3265766.

[13] "General Principles of the Curriculum, 2022 Revised Curriculum in Korea," Ministry of Education Notice No. 2022-33, 2022. [Online]. Available: https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N.

[14] J. Kim et al., "A study on the development of digital literacy education guidelines linked to the curriculum (translated by the author)," KERIS, CR 2023-1, 2023. [Online]. Available: https://www.keris.or.kr/main/ad/pblcte/selectPblcteRRInfo.do?mi=1138&pblcteSeq=13665.

[15] "Primary Curriculum, 2022 Revised Curriculum in Korea," Ministry of Education Notice No. 2022-33, Supplement 2, 2022. [Online]. Available: https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N.

[16] "Middle School Curriculum, 2022 Revised Curriculum in Korea," Ministry of Education Notice No. 2022-33, Supplement 3, 2022. [Online]. Available: https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N.

[17] "High School Curriculum, 2022 Revised Curriculum in Korea," Ministry of Education Notice No. 2022-33, Supplement 4, 2022. [Online]. Available: https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N.

[18] S. Elo and H. Kyngäs, "The qualitative content analysis process," *Journal of Advanced Nursing*, vol. 62, no. 1, pp. 107–115, Mar. 2008, doi: 10.1111/j.1365-2648.2007.04569.x.

[19] C. Ridsdale, J. Rothwell, M. Smit, H. Ali-Hassan, M. Bliemel, D. Irvine, D. Kelley, S. Matwin, and B. Wuetherick, "Strategies and best practices for data literacy education," *Knowledge Synthesis Report*, 2015.

[20] H. Y. Kim, "Analysis of Data Literacy in the Core Curriculum to Improve Students' 4C Skills: Communication, Collaboration, Critical Thinking, and Creativity," *The Korean Association of General Education*, vol. 14, no. 6, pp. 147–159, Dec. 2020, doi: 10.46392/kjge.2020.14.6.147.

[21] J. Kim, "A Content Analysis of Research Data Management Training Programs at the University Libraries in North America: Focusing on Data Literacy Competencies," *Journal of the Korean Society for information Management*, vol. 35, no. 4, pp. 7–36, Dec. 2018.

[22] S. W. Han "A Study on Design of Data Literacy Model Based on Digital Humanities," *Journal of the Korean Society for information Management*, vol. 37, no. 1, pp. 179–195, Mar. 2020.

[23] S. Choi, J. H. Jung, and S. W. Jung, "Concept and procedures of qualitative content analysis," *Qualitative Inquiry*, vol. 2, no. 1, pp. 127–155, 2016.

[24] H. Yang, J. Kim, and W. Lee, "Analysis of computing competency framework for data science," *in Proceedings of KACE*, vol. 25, no. 1(A), pp. 89–91, 2021.

[25] S. Jun et al., "Development of artificial intelligence education content system of K-12 for artificial intelligence convergence education," *Journal of KACE*, vol. 26, no. 2, pp. 71–88, 2023.

[26] C. T. Force, "Computing curricula 2020: Paradigms for global computing education november 2020," *ACM, IEEE-CS,* 2020.

[27] D. R. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," Theory Into Practice, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi: 10.1207/s15430421tip4104_2.

[28] S. Carretero, R. Vuorikari, and Y. Punie, DigComp 2.1: The Digital Competence Framework for Citizens. With Eight Proficiency Levels and Examples of Use. Publications Office of the European Union, 2017.

[29] Computer Science Teachers Association (CSTA), *K-12 Computer Science Standards: Revised 2017*. ACM, 2017.

[30] K. Farrell and J. Watts, *Data Science: Educators' Guide to the National Progression Award*. The University of Edinburgh, 2020.

[31] K. Farrell and J. Watts, "Learners' Guide to the National Progression Award in Data Science," Oct. 2020.

[32] R. Vuorikari, S. Kluzer, and Y. Punie, DigComp 2.2: The Digital Competence Framework for Citizens - With New Examples of Knowledge, Skills and Attitudes, JRC Research Reports JRC128415, Joint Research Centre, 2022. [Online]. Available: https://ideas.repec.org/p/ipt/iptwpa/jrc128415.html.

[33] S. Kim, Y. Jeon and T. Kim, "A Study on Data Preprocessing Content Knowledges According to School Level for Artificial Intelligence Education", *The Journal of Korean Association of Computer Education*, vol. 24, no. 4, pp. 1-12, 2021.

[34] B. Kim, B. Kim, and J. Kim, "Development and Validation of Data Science Education Instructional Model," *Journal of The Korean Association of Information Education*, vol. 26, no. 5, pp. 417–425, Oct. 2022, doi: 10.14352/jkaie.2022.26.5.417.

[35] "K-12 AI curricula: A mapping of government-endorsed AI curricula," UNESCO, Paris, France, ED-2022/FLI-ICT/K-12 REV, 2022. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000380602.locale=en.

[36] A. Joo and M. R. Kim, “Effects of Discussion Classes Using Data Visualization Materials on Data Literacy of Elementary School Students,” *The Journal of Korean Association of Computer Education*, vol. 27, no. 2, pp. 37–47, Mar. 2024, doi: 10.32431/kace.2024.27.2.004.

[37] S. W. Kim *et al.*, “Development of Modular Artificial Intelligence Education Textbooks to Cultivate Pre-service Teachers' Artificial Intelligence Literacy,” *The Journal of Korean Association of Computer Education*, vol. 26, no. 5, pp. 43-56, Sep. 2023.

[38] H. Heo and S. C. Kang, "Teacher competencies for designing artificial intelligence-integrated education," *The Journal of Korean Association of Computer Education*, vol. 26, no. 2, pp. 89-100, Mar. 2023.

[39] K. H. Kim *et al.*, “Development of Artificial Intelligence Convergence Education Common Curriculum for Graduate School and Verification of Teaching Efficacy Effectiveness,” *The Journal of Korean Association of Computer Education*, vol. 24, no. 3, pp. 89-98, May 2021.

[40] J. K. Kim and E. Sohn, “Development of Data Science Course and Analysis of Computational Thinking Effect for non-Major Students,” *The Journal of Korean Association of Computer Education*, vol. 24, no. 3, pp. 23-31, May 2023.