



Development of a digital literacy measurement tool for middle and high school students in the context of scientific practice

Mihyun Son¹ · Minsu Ha²

Received: 16 June 2024 / Accepted: 12 August 2024
© The Author(s) 2024

Abstract

Digital literacy is essential for scientific literacy in a digital world. Although the NGSS Practices include many activities that require digital literacy, most studies have examined digital literacy from a generic perspective rather than a curricular context. This study aimed to develop a self-report tool to measure elements of digital literacy among middle and high school students in the context of science practice. Using Messick's validity framework, Rasch analysis was conducted to ensure the tool's validity. Initial items were developed from the NGSS, KSES, and other countries' curricula and related research literature. The final 38 items were expertly reviewed by scientists and applied to 1194 students for statistical analysis. The results indicated that the tool could be divided into five dimensions of digital literacy in the context of science practice: collecting and recording data, analyzing and interpreting (statistics), analyzing and interpreting (tools), generating conclusions, and sharing and presenting. Item fit and reliability were analyzed. The study found that most items did not show significant gender or school level differences, but scores increased with grade level. Boys tended to perform better than girls, and this difference did not change with grade level. Analysis and Interpretation (Tools) showed the largest differences across school levels. The developed measurement tool suggests that digital literacy in the context of science practice is distinct from generic digital literacy, requiring a multi-contextual approach to teaching. Furthermore, the gender gap was evident in all areas and did not decrease with higher school levels, particularly in STEM-related items like math and computational languages, indicating a need for focused education for girls. The tool developed in this study can serve as a baseline for teachers to identify students' levels and for students to set learning goals. It provides information on how digital literacy can be taught within a curricular context.

Keywords Digital literacy · Scientific practice · Measurement tool · Middle and high school students

Extended author information available on the last page of the article

1 Introduction

Fostering scientific literacy is one of the most important goals of science education, and scientific literacy has been defined in various ways. Scientific literacy is sometimes described as the ability to use evidence and data to evaluate the quality of scientific information and claims presented by scientists and the media (NRC, 1996), or as the ability to understand and make decisions about changes in the world by drawing evidence-based conclusions using scientific knowledge (AAAS, 1993). The Next Generation Science Standards (NGSS), which can be considered a representative guideline for the direction of science education in the United States, also mentions scientific literacy in terms of students' understanding of scientific concepts, evaluation of scientific claims based on evidence, and participation in scientific practices (NRC, 2013). In this way, scientific literacy is the most fundamental competency for understanding the world and for continuous scientific engagement. In this context, digital literacy is essential for fostering scientific literacy in the digital world (Demirbag & Bahcivan, 2021; Mason et al., 2010; Walraven et al., 2009).

Korean science education research also emphasizes digital literacy in the context of scientific practices. The Korean Science Education Standards (KSES) divides scientific literacy into three dimensions: competences dimension, knowledge dimension, and participation & action dimension. Among these, the competences dimension includes scientific inquiry ability, scientific thinking ability, communication and collaboration ability, information processing and decision-making ability, and lifelong learning ability in a hyper-connected society (MOE et al. 2019). These five areas encompass both the skills traditionally emphasized in science education and those anticipated to be necessary in the future society characterized by the digital revolution. For instance, within the scientific inquiry ability, there are skills such as data transformation, engineering design and creation, and explanation generation and argumentation. Additionally, scientific thinking ability includes mathematical and computational thinking, while communication and collaboration ability includes the ability to express ideas. The 'information processing and decision-making ability' within the competences dimension involves the ability to search for, select, produce, and evaluate information and data. The emphasis on the importance of digital literacy and its integration with subject education can also be found in the curricula of various countries such as Singapore and Europe, as well as in reports from organizations like the OECD (Ei & Soon, 2021; Erstad et al., 2021; Polizzi, 2020).

The trend of science education reform is calling for changes in the relationship between scientific knowledge and scientific methods in science learning (Kawasaki & Sandoval, 2020). First, when students handle actual data, learning experiences related to data utilization skills can occur, ultimately aiming to cultivate scientific thinking and problem-solving abilities. The actual data used by students can take various forms, such as data collected by students in inquiry projects, searches in online data repositories, illustrations and tables in textbooks, or scientific publications (Hug & McNeill, 2008; Kerlin et al., 2010). Students need to

select appropriate data from these sources, classify it according to their objectives, and develop skills in collecting, storing, representing, and sharing data. In this process, they should be able to engage in activities such as data analysis and interpretation, utilizing mathematical and computational thinking, and participating in evidence-based arguments (NGSS Lead States, 2013; NRC, 2013).

Additionally, basic computational thinking is necessary to understand and solve socio-scientific issues related to real life. This requires the ability to use algorithmic thinking, data analysis and representation for modeling thinking, and simulation tools (Rodríguez-Becerra et al., 2020). The importance of computers in scientific inquiry has grown due to advancements in artificial intelligence, software platforms, and sensors. While there have been limitations in science education due to the lack of various data sets, the proliferation of sensors has made personalized data collection possible, facilitating the collection of data relevant to scientific inquiry contexts. Furthermore, the establishment of platforms for data sharing and environments that facilitate data analysis and visualization have made computer and digital-based scientific inquiry representative activities of scientific practice.

Digital literacy refers not only to the basic skills related to using digital devices but also to the complex skills that support learners by enhancing learning outcomes in digital environments. These skills include cognitive, social, and emotional skills (Eshet-Alkalai & Soffer, 2012). The meaning of digital literacy has expanded to include communication and content production using information and communication technology (ICT) (Mason et al., 2018), information retrieval and processing through new technologies (Siddiq et al., 2016), and communication with communities (da Silva & Heaton, 2017). Various countries and research organizations have presented diverse aspects of data literacy, which commonly include three main elements: 1) information and data, 2) communication and collaboration, and 3) technical skills (Bravo et al., 2021). These three elements commonly included in digital literacy largely overlap with the components of scientific literacy, indicating that digital literacy can be integrated with subject-specific digital competence education (Kotzebue et al., 2021).

Based on the relationship between these two literacies, many scholars have continued efforts to understand scientific literacy through digital literacy (Bliss, 2019; Da Silva & Heaton, 2017; Holincheck et al., 2022; Mardiani et al., 2024). They have introduced terms such as digital scientific literacy (Holincheck et al., 2022), aimed to develop critical evaluation skills for digital scientific materials (Bliss, 2019; Holincheck et al., 2022), engaged in inquiry activities using digital scientific materials (Mardiani et al., 2024), and examined the impact of information or data sharing—a component of digital literacy—on students' construction of scientific knowledge (Dewi et al., 2021; Mardiani et al., 2024). However, the evaluation tools used to assess the effectiveness of education have mostly focused on separately verifying digital literacy and subject content. Given that digital literacy includes both generic and subject-specific aspects (D-EDK, 2014; Kotzebue et al., 2021), most measurements have emphasized the generic part of digital literacy.

Studies aimed at developing digital literacy assessment tools have also emphasized the cross-curricular aspects of digital literacy, often constructing items in the form of exam questions (ACARA, 2018; Chetty et al., 2018; Covello & Lei,

2010; Jin et al., 2020), which makes it difficult for students to develop metacognitive understanding of the level of digital literacy they need to attain. Additionally, most tools are designed for use at a specific school level or age group (Cote & Milliner, 2016; Jin et al., 2020; Oh et al., 2021), making it challenging to longitudinally track changes in students' literacy levels.

Another aspect of this study is the evaluation tools for scientific literacy (or skills), which face challenges in finding forms that are applicable in the digital age. While traditional scientific literacy competencies have emphasized data analysis, representation, and sharing, there are difficulties in adapting these tools for the digital era. For instance, in the study by Gormally et al. (2012) on developing scientific literacy assessment tools, it is noted that students should have the basic scientific literacy to approach scientific phenomena quantitatively and possess various skills to apply this to problem-solving in everyday life (NRC, 2003). However, traditional tools derived from scientific inquiry and scientific methods carry inherent limitations. These tools often fail to accurately explain what is important in science, seem to perform inquiries only to explain theories (Osborne, 2014), and do not focus on activities (Ford, 2015). Consequently, to solve everyday problems in the digital world, there is a need for a new term that can encompass a broader meaning and have a sustained and widespread impact on our lives (Da Silva & Heaton, 2017).

Thus, the term 'Practice' is being used in place of scientific method or inquiry to represent the educational goal of teaching students how to reason and act scientifically in an integrated digital world (Osborne, 2014; Ford, 2015). Based on this discussion, we aim to develop a self-report measurement tool that can be utilized in classrooms, grounded in the important elements of digital literacy within the context of scientific practice. The specific research questions of this study are as follows:

RQ1. What is the content validity of the digital literacy assessment tool in the context of scientific practice?

RQ2. What validity evidence is identified in the statistical tests using evaluation data for the digital literacy assessment tool in the context of scientific practice?

RQ3. Are there significant gender and school level differences in the scores of the digital literacy assessment tool in the context of scientific practice?

2 Research method

The central research question of this study is to develop a digital literacy assessment tool based on strong validity evidence. Our RQ1 concerns the content validity of the developed assessment tool. Additionally, RQ2 involves collecting validity evidence through statistical methods using actual student data. Furthermore, RQ3 is a study on the application of the developed assessment tool. To verify the validity of the assessment tool developed in this study, Messick's (1995) validity framework was used. Messick (1995) defined validity as "an integrated judgment of the degree to which theoretical and empirical evidence supports the adequacy and appropriateness of interpretations and actions based on test scores." He proposed six aspects of validity: content, substantive, structural, generalizability, external, and consequential

(Messick, 1995). In this study, among Messick's six validity frames, content-based validity and substantive validity were verified using qualitative methods through the evaluation of scientists and the analysis of the student survey process during the item development process. The sections 'Initial Development through Literature Review' and 'Completion through Surveys with Scientists' pertain to content validity and correspond to RQ1. Subsequently, the sections 'Participants, Data Collection, and Analysis' correspond to RQ2 and RQ3.

2.1 Initial development through literature review

This study develops self-report items that measure digital literacy related to the scientific practice process. The goal is to present the functional objectives of digital-related scientific inquiry and develop items to identify the current level of students. Since this study defines the necessary skills for middle and high school students according to the 'inquiry process,' it uses the 'science and engineering practices standards' from NGSS and Korea's KSES as the basic framework. Additionally, it incorporates the difficulties and required student skills identified in various studies that combine scientific inquiry contexts with digital literacy. The digital-related inquiry process centers on activities beginning with data collection and analysis, followed by constructing and sharing conclusions. Ultimately, the items were developed in a four-stage structure: data collection, data analysis, drawing conclusions, and sharing. To emphasize the social practice of science learning in the digital age, the process of sharing was included, replacing the term 'communication' from NGSS's Practice with 'sharing (communication)' to reflect the importance of information sharing in the digital era (Elliott & McKaughan, 2014).

When examining the eight practices of the NGSS in the United States, terms that did not appear in the general scientific inquiry process are directly mentioned (NRC, 2013). Terms such as "Developing and using models," "Using mathematics and computational thinking," and "Constructing explanations and designing solutions" highlight the need to focus on these functions in scientific inquiry as science, engineering, and technology become increasingly integrated. Similarly, South Korea has developed and announced science education standards for future generations from 2014 to 2019. The KSES includes not only traditional scientific competencies and skills but also those anticipated to be necessary in a future society characterized by the digital revolution (Song et al., 2019). Additionally, the data literacy presented in the OECD 2030 report served as an important basis for item development (OECD, 2019). Many countries have recently set data literacy and digital literacy as goals within their educational curricula, and related research has been utilized in item development (Ei & Soon, 2021; Erstad et al., 2021; Polizzi, 2020). Therefore, by referencing research articles on scientific inquiry published between 2018 and 2022 that implemented programs related to cultivating competencies in data literacy or digital literacy or presented specific inquiry processes, the necessary skills were added (Aksit & Wiebe, 2020; Arastoopour Irgens et al., 2020; Chen et al., 2022; Clark et al., 2019; Gibson & Mourad, 2018; Kjelvik & Schultheis, 2019; Lichti et al., 2021; Son et al., 2018; Tsybulsky & Sinai, 2022; Wolff et al., 2019).

The tool developed in this study is a self-report measurement tool. Self-report tools in competency assessment can have limitations due to biases such as over-confidence (Moore & Healy, 2008). However, this tool is not intended to quantify abilities but rather to be used for learning assessments, allowing students to evaluate their own state and goals and reflect metacognitively. Our goal is for the developed assessment tool to be widely used in digital-based science classes conducted in schools. Therefore, the assessment tool was developed to include a Likert scale for self-reporting. Through this tool, students can evaluate their practical competencies in reflecting on themselves, as well as acquiring skills and knowledge (Demirbag & Bahçivan, 2021). It is about identifying their position in the learning goal achievement process and their ability to investigate and integrate additional information (Bråten et al., 2011). A self-report assessment tool can help students identify their current position and independently set future learning goals.

2.2 Completion through surveys with scientists

The 48 items completed through the literature review were sent to seven scientists researching advanced digital-based scientific fields to confirm the content validity of the items. Digital literacy in science is an essential scientific inquiry skill for students who will live in future societies and a necessary inquiry skill for high school students who plan to advance to STEM universities. However, as science and technology rapidly develop, the content and methods of education change accordingly, creating a time lag between the development of science and the development of science education. Therefore, to bridge this gap, it is necessary to review the opinions of scientists currently conducting research in relevant fields. A total of seven scientists reviewed these items, each with more than 10 years of research experience and actively engaged in recent research activities (see Table 1). The scientists confirmed the content validity of each item and, when modifications were necessary, described the reasons and directions for the revisions.

After undergoing content validity evaluation, the final 48 items were administered to 43 middle school students to verify substantive aspect of construct validity. This process aimed to confirm whether students could understand the content of the items and respond as intended. It was checked if the terms were appropriate for the students' cognitive level and whether the questions were understood as intended. During

Table 1 Information on the science expert who reviewed the content validity

Number	Research Area	Research Experience	Gender
1	Bio Science	10–20 years	Female
2	Computer Science	Over 20 years	Male
3	Astronomy	10–20 years	Male
4	Electrical Engineering	Over 20 years	Male
5	Life Science	Over 20 years	Female
6	Nano Science	Over 20 years	Male
7	Optics	Over 20 years	Male

this process, some students had difficulty interpreting certain items, so examples were added, or the items were revised into language easier for students to understand. The survey took approximately 30 min, and it was confirmed that students were able to focus better on the survey when guided by a supervising teacher. The final revised items were confirmed, and a large amount of data was collected from middle and high school students for statistical validity verification.

2.3 Participants, data collection and analysis

To verify statistical validity, the finalized items were administered to over a thousand students. A total of 1,194 students participated, including 651 middle school students and 543 high school students. The survey was conducted in five schools: one middle school and one high school located in a major city, and one middle school and two high schools located in a small city. Regarding the gender of the participants, there were 537 male students (331 middle school students) and 657 female students (320 middle school students). To minimize data bias related to educational level and gender, participants were recruited considering various regions and a balanced gender ratio. This study involved minors as vulnerable participants, and the entire process was conducted with approval from the IRB of the relevant research institution before the study commenced.

Using data from over a thousand students, statistical tests were conducted to confirm item fit, reliability, differential item functioning, criterion-related validity, and structural validity. The statistical tests were performed using item response theory-based analyses, such as Rasch analysis, suitable for Messick's validity framework (Wolfe & Smith, 2007). In the Rasch analysis, item fit was checked using Infit MNSQ and Outfit MNSQ, with the criterion value set between 0.5 and 1.5 (Boone et al., 2014). Person reliability and item reliability were verified using Rasch analysis. To confirm construct validity based on internal structure, dimensionality was tested in Rasch analysis to satisfy unidimensionality (Boone et al., 2014). For external validity, five additional self-report items measuring core competencies in Korean science subjects were included in the field test alongside the developed items. These self-report items for measuring core competencies in science subjects had been previously field-tested on more than 2000 Korean adolescents and were known for their high validity and reliability (Ha et al., 2018). Additionally, since these core competency items included some scientific inquiry skills such as information processing, data transformation, and analysis, they were appropriate for securing external validity. Lastly, group score comparisons were conducted to identify any gender or school level differences in the scores of the developed tool. Rasch analysis was performed using Winsteps 4.1.0, and all other statistics were analyzed using SPSS 26.

3 Research results

3.1 RQ1: Content validity of items as judged by scientists

These are the results of the scientist evaluation to verify the internal validity of the developed items. The scientists agreed that, while science inquiry education in

schools is generally well-conducted, there is a need for changes in its approach. The scientists reviewed the items and assessed the content validity regarding whether each skill was necessary for middle and high school students. We analyzed the Content Validity Index (CVI) using their evaluations. The acceptability of the CVI value depends on the number of panelists; since there were seven scientists in this study, a CVI of 0.83 or higher is required for acceptability (Lynn, 1986). Most items had values of 0.86 or higher, but a few items had lower values. The seven items out of the total 48 that did not meet the acceptable range are as follows (Table 2).

Generally, the items included in the analysis and interpretation process had lower content validity, whereas items related to data collection, recording, drawing conclusions, and sharing processes had overall high content validity. Analyzing the items with low content validity reveals two main points. First, students showed negative opinions regarding expressing scientific discovery results using mathematical models or formulas. Second, while understanding and utilizing pre-developed or pre-written computer programs or code is considered a necessary skill, students did not see the need for a deep understanding required to develop or debug programs themselves.

The scientists mentioned that the reason they did not consider these functions important is that there should be a distinction between students who will major in science in university and those who need general scientific literacy. They thought it unnecessary to practice creating mathematical models in general science education, as it might not be important or possible depending on the type of scientific inquiry. Furthermore, they were concerned that overly generalizing results to fit into mathematical models at the students' level of mathematics might lead to misconceptions. Regarding learning computer programming skills, they were apprehensive about the potential focus on programming languages themselves. Since programming languages and software continually evolve, they believed there was no need to become familiar with the grammar of computer languages. Instead, they emphasized the importance of analyzing how to process problems and predicting the outcomes of those processes. Based on expert review, six items deemed more appropriate for university-level science majors were deleted from the study. Additionally, four items with overlapping content were combined into more comprehensive questions, resulting in a final set of 38 items.

Table 2 Items rated low in content validity by science experts

Construct	Items	CVI
Analysis and interpretation	I know what to do when an error occurs in the software used for data analysis	0.71
	I can find patterns in complex results using computer software	0.71
	I can use programs like Excel or Scratch to measure and analyze data	0.57
	I can use programs like Python to measure and analyze data	0.57
	I can use various programs like Matlab, Python, and Excel to measure data and predict unexpected values	0.57
Conclusion generation	I can interpret data and create logical arguments based on it	0.71

3.2 RQ2: Validity evaluation based on statistics

The final set of items was administered to 1,199 students, and the collected data was analyzed to verify validity through various methods. The first analysis conducted was dimensionality analysis. We categorized the digital competencies in the context of scientific practice into four dimensions: data collection and recording, analysis and interpretation, conclusion generation, and sharing and presentation. We composed various items for each factor. Each item was intended to contribute to measuring its respective construct, and each factor was assumed to be unidimensional. If multiple items for a specific construct do not assume unidimensionality and are instead divided into multiple components internally, they are not valid from a measurement perspective.

We performed PCA analysis using residuals from Rasch analysis for this evaluation (Table 3). If there are consistent patterns in the parts of the data that do not align with the Rasch measurement values, it suggests the presence of an unexpected dimension. According to Bond et al. (2020), if the Eigenvalue of the unexplained variance exceeds 2, there is a possibility of another dimension, while if it is below 2, the construct can be assumed to be unidimensional. As shown in Table 3, the first unexplained variance for data collection and recording, conclusion generation, and sharing and presentation does not exceed 2. However, for the analysis and interpretation items, the first unexplained variance is 2.555, which significantly exceeds 2. We further conducted an exploratory factor analysis for this construct and found that splitting it into two dimensions—items 1 to 8 and items 9 to 12—meets the unidimensionality assumption. Upon close examination, we discovered that items 1 to 8 pertain to the analysis and interpretation of statistical data and graphs, while items 9 to 12 pertain to the use of analytical tools, indicating a difference in content (see Appendix). Therefore, we concluded that it is more valid to separate this part into two dimensions. Consequently, the valid use of this assessment tool is determined to be the analysis of five categories: data collection and recording, analysis and interpretation 1 (statistics), analysis and interpretation 2 (analytical tools), conclusion generation, and sharing and presentation.

Item fit refers to information about whether there are any unusual respondent reactions to specific items. For example, if a significantly higher number

Table 3 Results of dimensionality analysis using principal component analysis of the Rasch model

	Raw variance explained by			Raw unexplained variance in			
	Measures	Persons	Items	1st contrast	2nd contrast	3rd contrast	
Data collection and recording	10.286	6.791	3.494	1.581	1.394	1.277	
Analysis and interpretation	All	12.466	7.873	4.593	2.555	1.415	1.217
	1–8	10.364	7.795	2.569	1.623	1.238	
	9–12	6.512	5.213	1.299	1.541	1.284	1.178
Conclusion generation	13.830	11.564	2.266	1.434	1.258	1.235	
Sharing and presentation	11.471	9.287	2.184	1.498	1.413	1.208	

of respondents agree or disagree with a particular item compared to other items, the item fit decreases. In Rasch analysis, item fit is checked using Mean Square (MNSQ). Rasch analysis also allows for checking various types of reliability. Person reliability (PR) checks how reliably items measure the respondent's abilities, while item reliability (IR) checks how appropriate the respondent's abilities are for verifying the quality of the items. Additionally, internal consistency reliability is verified using Cronbach's alpha (CA). To see if a specific item supports or hinders the internal consistency of the construct, the Cronbach alpha if the item is deleted (Alpha if item deleted, AIC) is also checked. We recorded all these results in a single table. The comprehensive information in the table reveals the following (Table 4).

Overall, all items have adequate fit. The person reliability and item reliability identified in the Rasch analysis both exceeded or approached 0.8 or 0.9, indicating very high reliability. The internal consistency reliability of the items also exceeded 0.8, showing excellent reliability. Additionally, no items were found to significantly affect internal consistency reliability. Based on item fit and reliability information, we can conclude that there are no particular issues that need to be addressed in the developed items.

The following validity evidence pertains to generalizability validity (Table 5). Using the measurement values related to digital competence, score comparisons were conducted across various groups such as gender and grade levels. The premise for comparing scores between groups is that the measurement tool functions equally across different groups. Evidence regarding generalizability validity can be confirmed through differential item functioning (DIF) analysis. In Rasch analysis, DIF is checked using the difference in DIF values (DIF C), Rasch-Welch t-test, and Mantel chi-square test. The table presents DIF C (DIF contrast), the significance of the Rasch-Welch t-test (RW p), and the significance of the Mantel chi-square test (MC p).

Regarding the interpretation of DIF differences, a value between 0.43 and 0.64 indicates a moderate level of DIF difference, while a value exceeding 0.64 indicates a large DIF difference (Zwick et al., 1999). Although there were no items exceeding 0.64, one item showed a DIF difference exceeding 0.43 for gender, and one item showed a similar difference for grade levels. When using the significance values of the Rasch-Welch t-test and Mantel chi-square test, more items were found to have a p -value of 0.00. For gender, five items showed a p -value of 0.00, and for grade levels, about eight items showed similar results. We concluded that some items in the digital competence tool exhibit differential item functioning. This may be due to the inconsistent application of various elements within the items across groups. For example, the ability to understand graphs and tables in item 7 of the analysis and interpretation section showed DIF for both gender and grade level, indicating that this item functions differently across these groups. Nonetheless, considering that the overall DIF differences are not large and that experiences related to digital competence may vary significantly by gender and grade level, it can be interpreted that no severe DIF was found in the items.

We also examined criterion-related validity. The scores for science-related digital competence are closely related to core science competencies and interest in science or information and computer subjects (Table 6). Therefore, the scores of our

Table 4 Item fit and reliability using Rasch analysis and internal consistency reliability (Cronbach alpha)

	Item	Mea- sure	Infit		Outfit		AIC
			MNSQ	ZSTD	MNSQ	ZSTD	
Data collection and recording PR: 0.88 IR: 0.98 CA: 0.92	1	-0.22	1.04	0.86	1.03	0.63	0.91
	2	-0.34	1.24	5.28	1.22	4.78	0.91
	3	-0.13	0.77	-5.74	0.78	-5.37	0.91
	4	0.71	1.32	6.90	1.34	7.23	0.92
	5	-0.16	1.17	3.74	1.15	3.29	0.91
	6	-0.31	0.75	-6.43	0.75	-6.34	0.91
	7	0.30	0.78	-5.44	0.79	-5.25	0.91
	8	0.06	1.09	1.97	1.09	2.13	0.91
	9	0.23	0.89	-2.55	0.91	-2.18	0.91
	10	-0.14	0.87	-3.08	0.87	-3.08	0.91
Analysis and interpretation 1 (statistics) PR: 0.88 IR: 0.98 CA: 0.92	1	0.24	1.06	1.33	1.05	1.10	0.91
	2	-0.16	1.01	0.15	0.99	-0.10	0.91
	3	0.17	1.11	2.35	1.11	2.47	0.91
	4	0.47	1.04	1.00	1.05	1.17	0.91
	5	0.25	0.86	-3.39	0.85	-3.47	0.91
	6	0.24	0.82	-4.41	0.82	-4.22	0.91
	7	-0.77	1.02	0.59	1.02	0.48	0.91
	8	-0.43	1.02	0.48	0.99	-0.19	0.91
Analysis and interpretation 2 (analytical tools) PR: 0.83 IR: 0.98 CA: 0.86	9	-0.18	1.06	1.37	1.04	0.92	0.83
	10	-0.12	0.86	-3.44	0.85	-3.61	0.81
	11	-0.41	1.03	0.71	1.01	0.25	0.83
	12	0.71	1.01	0.30	1.02	0.38	0.83
Conclusion generation PR: 0.91 IR: 0.96 CA: 0.94	1	-0.22	0.86	-3.16	0.84	-3.50	0.93
	2	0.03	0.91	-2.03	0.87	-2.92	0.93
	3	-0.16	0.92	-1.80	0.88	-2.53	0.93
	4	-0.30	0.91	-1.90	0.87	-2.86	0.93
	5	0.72	1.36	7.36	1.36	7.09	0.94
	6	0.09	0.98	-0.50	0.94	-1.32	0.93
	7	0.15	1.01	0.15	0.97	-0.61	0.93
	8	-0.31	0.97	-0.75	0.92	-1.63	0.93
Sharing and presentation PR: 0.89 IR: 0.91 CA: 0.93	1	0.08	1.02	0.51	1.01	0.19	0.92
	2	-0.16	1.00	-0.05	0.97	-0.65	0.92
	3	-0.27	0.86	-3.31	0.83	-3.90	0.92
	4	-0.12	0.92	-1.76	0.90	-2.25	0.92
	5	0.11	0.99	-0.19	0.96	-0.85	0.92
	6	-0.07	0.81	-4.59	0.79	-4.91	0.92
	7	0.05	1.15	3.23	1.13	2.78	0.93
	8	0.38	1.18	3.89	1.18	3.72	0.93

Table 5 Results of differential item functioning using Rasch analysis

	Item	Gender			School		
		DIF C	RW p	MC p	DIF C	RW p	MC p
Data collection and recording	cd1	0.14	0.13	0.24	-0.07	0.49	0.58
	cd2	0.29	0.00	0.01	0.12	0.22	0.10
	cd3	0.08	0.39	0.61	0.00	1.00	0.47
	cd4	-0.18	0.05	0.13	0.35	0.00	0.02
	cd5	0.00	1.00	0.89	0.17	0.07	0.05
	cd6	0.00	1.00	0.91	-0.25	0.01	0.01
	cd7	0.00	1.00	0.95	-0.07	0.44	0.14
	cd8	-0.15	0.11	0.08	-0.25	0.01	0.02
	cd9	-0.10	0.28	0.52	0.12	0.21	0.48
	cd10	-0.02	0.81	0.78	-0.15	0.10	0.09
Analysis and interpretation 1 (statistics)	ad1	-0.21	0.04	0.03	0.09	0.41	0.50
	ad2	-0.19	0.07	0.08	-0.17	0.11	0.15
	ad3	-0.11	0.29	0.40	0.38	0.00	0.00
	ad4	-0.03	0.80	0.59	0.20	0.06	0.22
	ad5	-0.09	0.39	0.69	0.13	0.22	0.28
	ad6	0.00	1.00	0.50	0.17	0.11	0.18
	ad7	0.49	0.00	0.00	-0.46	0.00	0.00
	ad8	0.19	0.07	0.07	-0.37	0.00	0.00
Analysis and interpretation 2 (analytical tools)	ad9	0.15	0.15	0.20	0.37	0.00	0.00
	ad10	0.21	0.05	0.02	0.06	0.56	0.34
	ad11	-0.03	0.80	0.66	-0.40	0.00	0.00
	ad12	-0.32	0.00	0.00	0.00	1.00	0.31
Conclusion generation	cl1	-0.10	0.39	0.36	0.30	0.01	0.01
	cl2	0.39	0.00	0.00	0.12	0.31	0.39
	cl3	0.00	1.00	0.83	-0.16	0.18	0.16
	cl4	0.14	0.23	0.12	-0.11	0.34	0.48
	cl5	-0.28	0.02	0.02	0.15	0.19	0.33
	cl6	-0.09	0.44	0.32	-0.26	0.03	0.02
	cl7	-0.14	0.25	0.23	0.32	0.01	0.02
	cl8	0.08	0.51	0.42	-0.37	0.00	0.01
Sharing and presentation	si1	-0.33	0.00	0.01	0.17	0.12	0.08
	si2	0.00	1.00	0.68	0.28	0.01	0.03
	si3	0.08	0.44	0.50	-0.06	0.60	0.79
	si4	-0.07	0.50	0.39	0.14	0.21	0.11
	si5	-0.15	0.15	0.21	0.14	0.21	0.23
	si6	0.11	0.31	0.35	-0.07	0.51	0.48
	si7	0.14	0.19	0.20	-0.42	0.00	0.00
	si8	0.23	0.03	0.03	-0.17	0.12	0.09

Table 6 Correlation between scores by component factors and external criterion scores

	Science core competencies*	Interest in science subjects**	Interest in computer/information subjects**
Data collection and recording	0.726 [‡]	0.444 [‡]	0.423 [‡]
Analysis and interpretation 1	0.726 [‡]	0.463 [‡]	0.349 [‡]
Analysis and interpretation 2	0.599 [‡]	0.349 [‡]	0.422 [‡]
Conclusion generation	0.760 [‡]	0.449 [‡]	0.329 [‡]
Sharing and presentation	0.765 [‡]	0.436 [‡]	0.361 [‡]

*Pearson's r **Spearman's ρ ‡ $p < 0.001$

developed science digital competence should show significant correlations with general science core competency scores and interest in science and computer subjects. To verify this, we conducted a correlation analysis. We selected five items developed by Ha et al. (2018) to generate scores for science core competencies. We also collected Likert scale scores for the items "Do you like science?" and "Do you like computer or information subjects?". The correlations between the five variables we developed and the three external criteria (science core competencies, interest in science subjects, and interest in computer/information subjects) are presented in Table 6. Since interest in subjects was collected using single Likert scale items, Spearman's rho correlation coefficients were used for analysis, while Pearson's correlation coefficients were used for the others.

Science digital competence showed a high correlation with science core competency scores. All correlations were significant at the 0.001 level, with r values exceeding 0.7, indicating a very strong correlation. There were also significant correlations at the 0.001 level with interest in science subjects and computer/information subjects. These results confirm that our developed science digital competence assessment tool is related to other similar indicators and operates as a valid measurement tool.

3.3 RQ3: Gender and school level differences in the scores of the digital literacy assessment tool

Our final statistical analysis concerns whether there are score differences in the assessment tool we developed based on gender and grade level. As discussed in the introduction, it is known that both science and digital competence have gender effects, with males generally showing higher competence or interest (Divya & Haneefa, 2018; Esteve-Mon et al., 2020; Gebhardt et al., 2019). Additionally, as students progress to higher school grades, their learning in science digital competence is expected to improve, resulting in higher competence scores. To confirm if our data exhibited these trends, we conducted a two-way ANOVA and presented the results in graphs and tables (Fig. 1 and Table 7). The graphs show the mean scores and standard errors for each group to provide an intuitive comparison of overall scores.

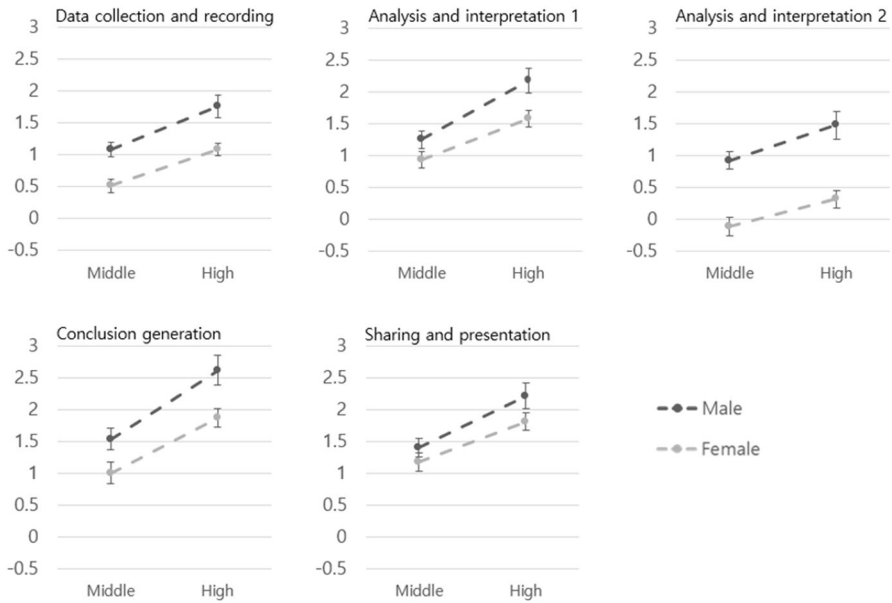


Fig. 1 Mean and standard error of scores by gender and school level

Table 7 Results of two-way ANOVA by gender and school level

	gender		school		gender * school	
	F	η^2	F	η^2	F	η^2
Data collection and recording	27.04 [‡]	0.02	27.12 [‡]	0.02	0.17 ^{ns}	0.00
Analysis and interpretation 1	9.70 [‡]	0.01	29.04 [‡]	0.02	0.93 ^{ns}	0.00
Analysis and interpretation 2	48.42 [‡]	0.04	9.47 [‡]	0.01	0.17 ^{ns}	0.00
Conclusion generation	13.28 [‡]	0.01	30.18 [‡]	0.02	0.38 ^{ns}	0.00
Sharing and presentation	4.06 [†]	0.00	22.00 [‡]	0.02	0.34 ^{ns}	0.00

[‡] $p < 0.001$, [†] $p < 0.01$

The key statistical results of the two-way ANOVA, including F -values, significance levels, and effect sizes, are summarized in a Table 7.

Examining the scores for the five items across four groups divided by gender and grade level, we observed consistent trends across all areas. For the five items of science digital competence, male students scored higher than female students, and high school students scored higher than middle school students. Notably, the most significant gender effect size was observed in the analysis and interpretation 2 category. Unlike analysis and interpretation 1, analysis and interpretation 2 involves the use of mathematical tools, computer coding, and programming languages like Python. This suggests that male students had significantly more experience and learning related to these areas compared to female students.

4 Discussions

4.1 RQ1. The content validity of the digital literacy assessment tool in the context of scientific practice

The purpose of this study was to develop a valid assessment tool to evaluate the level of digital literacy in the context of scientific practice for middle and high school students and to establish indicators of digital literacy in scientific practice. To this end, we developed the initial items through literature review and expert Delphi surveys, applied them to middle and high school students to verify statistical validity, and investigated whether the items could be applied regardless of gender and school level to finalize the items. Through this process, we identified a consensus on the elements and levels of digital literacy required in the context of scientific practice among scientists, national curricula, and empirical experiences in classroom settings. Additionally, considering that digital literacy is not merely the ability to use technology but also complements the enhancement of students' learning abilities in the context of science education (Yasa et al., 2023), we can propose specific directions for 'learning by doing' in science classes by providing empirical indicators of scientific practice and digital literacy.

Based on research from various countries and major institutions on specific scientific inquiry activities related to digital literacy, we initially developed 48 items. We then had scientists review whether each item was necessary for science majors or for general middle and high school students through two rounds of validation. Through this process and refinement, we finalized a total of 38 items. This process revealed differences between the digital literacy levels scientists believe students should have and the level of digital literacy needed for scientific inquiry performed in classroom settings. Scientists did not consider the criteria emphasizing complex skills, tool usage, or programming languages to be particularly important. They also expressed concerns that generalizations through formulas without sufficient theoretical background might lead to misconceptions. This indicates that the primary goal of science education, which is to develop students' thinking and problem-solving skills, remains unchanged. It also suggests the need for more detailed standards and application plans to avoid instrumentalism and ensure that the purpose of digital literacy aligns with the level students need to learn.

Digital competence in the context of scientific practice was divided into four dimensions: data collection and recording, analysis and interpretation, conclusion generation, and sharing and presentation, and dimensionality analysis was conducted. The dimensionality analysis revealed that the 'analysis and interpretation' part did not form a single dimension. An exploratory factor analysis showed that it split into statistical processing and the use of analytical tools. Thus, digital competence in the context of scientific practice was confirmed to be divided into five dimensions: data collection and recording, analysis and interpretation (statistics), analysis and interpretation (analytical tools), conclusion generation, and sharing and presentation.

Generally, digital literacy is theoretically composed of several dimensions, but empirical measurements of digital literacy often result in a single dimension or show strong correlations between elements (Aesaert et al., 2014; Demirbag and Bahcivan, 2021; Fraillon et al., 2019). While existing digital literacy developments encompass universal content, this study constructed elements within the context of scientific practice.

This indicates that when digital literacy education is conducted within the context of specific subjects, it is more likely that only certain elements, tailored to the characteristics of the subject, will be learned rather than all elements of digital literacy. So, it implies that digital literacy training tailored to specific subjects can facilitate the smooth operation of classes when teaching subjects that require digital literacy.

Furthermore, this implies that general digital literacy and digital literacy within specific subject contexts may differ. In the case of data literacy, which is similar to digital literacy, research has emphasized competencies within particular subject contexts, leading to the development of terms, definitions, and measurement tools such as scientific data literacy (Son & Jeong, 2020; Qiao et al., 2024; Qin and D'ignazio, 2010). However, there has been limited research on digital literacy within specific subject contexts. This study may serve as practical evidence supporting the argument that universal literacies, such as digital literacy and data literacy, require a different perspective on definition and measurement when learned within the context of specific subjects.

4.2 RQ2. Validity evidence identified in the statistical tests

The analysis of item fit and reliability showed that the item fit was generally appropriate across all items. The reliability of the items was measured using person reliability (PR), item reliability (IR), and Cronbach's alpha (CA), all of which were found to be above 0.8, indicating very high reliability. In addition to the content validity of the developed items, we examined criterion-related validity to confirm additional validity. Since the developed items pertain to digital competence in the context of scientific practice, it was assumed that scientific competence and interest in computers would be closely related to the results of these items. Therefore, additional survey questions on scientific competence and interest in computers and information were analyzed. The results showed significant correlations at the 0.001 level with both interest in science subjects and interest in computer/information subjects. Thus, we confirmed that the tool developed in this study operates validly.

Since we developed digital literacy items in the context of scientific practice for middle and high school students, it is necessary to confirm the generalizability across both school levels and between genders. We conducted DIF analysis to compare scores between groups, assuming that the measurement tool performs equally across different groups. The analysis showed that one item had a moderate difference by school level, and one item had a moderate difference by gender. Using the significance levels of the Rasch-Welch t-test and Mantel chi-square test, we found differences in five items by gender and eight items by school level. Gender

differences were evenly distributed across factors, while school level differences mostly occurred in the analysis and interpretation factors.

These items were related to mathematical knowledge and the use of computer languages, indicating that these competencies may vary as students' mathematical concepts and computer language skills increase (Fraillon et al., 2019). Lazonder et al. (2020) found that digital skills are influenced more by early exposure to digital tools than by age. However, higher-order thinking skills such as analysis and interpretation require not only early exposure but also cognitive level and understanding of subjects like mathematics, science, and computer science.

4.3 RQ3. Gender and school level differences in the scores of the digital literacy assessment tool

We conducted a two-way ANOVA to explore the differences by gender and grade level more deeply, confirming that digital and scientific literacy increase with higher grade levels. This trend has been confirmed by various studies (ACARA, 2018; Kim et al., 2019). When examining gender differences, we found that male students scored higher than female students across all items, with the most significant differences observed in items related to computer coding and software. The effect size was greater for male students, contrasting with the general trend where female students often score higher in science concept learning (Fraillon et al., 2019).

In our study, more items focused on functional aspects rather than conceptual ones, possibly giving male students an advantage in technical tasks (Divya & Haneefa, 2018; Esteve-Mon et al., 2020; Gebhardt et al., 2019). Additionally, many items were related to computers and mathematics, where male students tend to exhibit higher overconfidence (Adamecz-Völgy et al., 2023). The self-report nature of the survey may also have contributed to these results, as female students might underreport their abilities and confidence in STEM fields compared to their actual capabilities (Hand et al., 2017; Sobieraj & Krämer, 2019).

Consequently, students believe that their digital literacy within the context of scientific practices increases with age, and male students tend to rate themselves higher than female students across all categories. This suggests that male students find technical tasks easier and have reached a higher level, particularly in areas where mathematics and computer coding are integrated into scientific practices, compared to female students. Although this study is based on self-reported assessments, it can be inferred that there are actual differences in ability, not just in interest or confidence, among middle and high school students who have some understanding of their capabilities. These findings are consistent with previous research indicating that female students lag behind male students in STEM-related skills (Divya & Haneefa, 2018; Esteve-Mon et al., 2020; Fraillon et al., 2019; Gebhardt et al., 2019). Therefore, it is necessary to develop instructional strategies in science education to cultivate these competencies.

5 Conclusion and direction of future studies

In this study, we developed a measurement tool for digital literacy in the context of scientific practice for middle and high school students. Based on a literature review and a Delphi study with scientists, an initial draft was created and then applied to Korean middle and high school students. Through a statistical validation process, the tool was finalized. Assuming that digital competence should combine both general and subject-specific digital competencies, we aimed to establish specific criteria for digital literacy integrated with scientific practice. The developed items are applicable in both middle and high schools, with only a few items showing gender-related differences, which are not significant enough to limit their use.

Since the developed measurement tool consists of self-report items, it is important to consider the potential issues of overconfidence bias and the tendency to measure higher actual performance in digital literacy compared to conceptual understanding (Porat et al., 2018). However, this study is significant in that it approached digital literacy in a subject-specific context and presented an assessment tool with concrete and practical science lessons in mind to enhance digital competence. It can be universally used in various science subjects, providing guidance for teachers and students on the objectives of their participation in science classes. Understanding the characteristics of the various elements of digital literacy in the context of scientific practice can lead to the development of specific teaching and learning methods to enhance the corresponding competencies. This suggests that digital literacy, within the context of specific subjects, requires a different perspective in terms of its definition and measurement.

The items developed in this study are designed to be used in both middle and high schools, making them suitable for longitudinal research by other researchers. Given the technical changes and software developments, some items may need to be modified, and future related studies are expected to adapt these items accordingly. Additionally, it is necessary to more closely examine the reasons why female students have lower digital literacy, particularly in STEM-related fields, within the context of scientific practices compared to male students, and to explore strategies to reduce this gap.

Appendix

Data collection and recording

- 1 I know reliable websites and can search for appropriate papers or books when conducting theoretical research to solve questions (e.g., knowing which sites to access to find existing studies on the population of our neighborhood or air quality)
- 2 I can use drives (e.g., Google Drive or OneDrive) to store and effectively manage my data
- 3 I can identify variables according to questions and hypotheses and determine what data or information needs to be collected (e.g., considering and measuring factors like carbon dioxide, air quality, and tree types to know if trees help improve classroom air quality)
- 4 I know how to enter data into spreadsheets like Excel

- 5 I understand that sensors or measuring devices do not always measure accurate values
 - 6 I can think about what to consider to determine if the collected data is reliable
 - 7 I know what to consider when selecting sensors to measure data
 - 8 I know how to deal with errors in the devices I use
 - 9 I understand the difference between data being valid and data being reliable and can explain the meaning of each
 - 10 I try to devise and improve various methods of data collection
-

Analysis and interpretation 1 (statistics)

- 1 I attempt to analyze data in various ways (e.g., by day of the week, date, time, correlation with other variables)
 - 2 I can find recurring patterns when converting data into graphs (meaning abstraction)
 - 3 I can distinguish between correlation and causation (e.g., 'higher temperatures cause more photosynthesis' is causation, and 'students with higher math scores also have higher language scores' is correlation)
 - 4 I understand and can interpret the meanings of statistical results such as standard deviation, variance, mean, and maximum values
 - 5 When analyzing collected data, I can identify causes of potential errors and limitations in the data collection process to avoid excessive generalization (e.g., explaining why results from our classroom should not be generalized to all classrooms)
 - 6 While interpreting graphs, I can use scientific background knowledge to explain why certain inquiry results occurred
 - 7 I can understand and explain pictures, tables, and data
 - 8 I can compare and use different types of graphs and tables, understanding their characteristics and usage (e.g., knowing when to use pie charts, line graphs, or bar graphs)
-

Analysis and interpretation 2 (analytical tools)

- 1 I can use mathematical tools or techniques to calculate data (e.g., setting up Excel formulas for complex calculations)
 - 2 I can use computer languages for statistical analysis (block coding or text coding) (e.g., using Entry, Python, or Scratch to find mean, standard deviation, mode, etc.)
 - 3 I can understand the meaning of codes written by others in Entry or Scratch
 - 4 I can understand the meaning of codes written by others in Python or R
-

Conclusion generation

- 1 I can ethically consider and evaluate various information and alternatives during the problem-solving process
 - 2 I can thoroughly discuss the impacts of my solutions on other related fields
 - 3 I can derive creative conclusions by combining newly found information with what I already know
 - 4 I can synthesize various information to draw conclusions that help solve problems
 - 5 I can draw trend lines from current data to predict trends
 - 6 I can objectively evaluate the strengths and limitations of my conclusions
 - 7 I can explain how my conclusions are related to scientific and social issues
-

8 I can self-evaluate and revise the solutions I propose to solve problems

Sharing and presentation

- 1 I can make logical arguments using data or scientific results
 - 2 I can communicate with other students using presentations, online software, discussion boards, etc
 - 3 I can share solutions to problems with other students through shared documents or platforms like Google Docs or Padlet
 - 4 I know methods for sharing information or knowledge (e.g., Google Drive, writing shared documents)
 - 5 I can effectively present new information using computer programs (Excel, PowerPoint, Hangul)
 - 6 I know how to present information in a way that makes my written words, speeches, graphs, pictures, posters, etc., easily understood by others
 - 7 I can use various computer programs (video editing, photo editing, document creation, using formulas, or drawing graphs)
 - 8 I know how to present information beautifully to make it aesthetically pleasing to people
-

Acknowledgements This research was supported by the National Research Foundation of Korea (Grant Number NRF-700-20230072) and the 4th BK21 Infosphere Science Education Research Center granted by Ministry of Education of the Republic of Korea.

Author contributions Mihyun Son was responsible for the design of this study, data collection, data analysis, and writing the paper. Minsu Ha was responsible for data analysis and writing the paper.

Funding Open Access funding enabled and organized by Seoul National University. This research was supported by the National Research Foundation of Korea (Grant Number NRF-700-20230072) and the 4th BK21 Infosphere Science Education Research Center granted by Ministry of Education of the Republic of Korea.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamecz-Völgyi, A., Jerrim, J., Pingault, J. B., & Shure, D. (2023). Overconfident boys: The gender gap in mathematics self-assessment. IZA Discussion Paper No. 16180. <https://doi.org/10.2139/ssrn.4464593>

- Aesaert, K., Van Nijlen, D., Vanderlinde, R., & van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: Using item response theory for the development and validation of an ICT competence scale. *Computers & Education*, *76*, 168–181.
- Aksit, O., & Wiebe, E. N. (2020). Exploring force and motion concepts in middle grades using computational modeling: A classroom intervention study. *Journal of Science Education and Technology*, *29*(1), 65–82. <https://doi.org/10.1007/s10956-019-09800-z>
- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. Oxford University Press.
- Arastoopour Irgens, G., Dabholkar, S., Bain, C., Woods, P., Hall, K., Swanson, H., Horn, M., & Wilensky, U. (2020). Modeling and measuring high school students' computational thinking practices in science. *Journal of Science Education and Technology*, *29*(1), 137–161.
- Australian Curriculum, Assessment and Reporting Authority (ACARA). (2018). *National Assessment Program –ICT literacy years 6 & 10 2017 report*. Sydney, NSW: ACARA. Retrieved from: https://www.nap.edu.au/docs/default-source/default-documentlibrary/2017napictreport_final.pdf?sfvrsn=2.
- Bliss, A. C. (2019). Adult science-based learning: The intersection of digital, science, and information literacies. *Adult Learning*, *30*(3), 128–137.
- Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*(4th ed.). New York: Routledge.
- Boone, W., Staver, J., & Yale, M. (2014). *Rasch analysis in the human sciences*. Springer.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48–70.
- Bravo, M. C. M., Chalezquer, C. S., & Serrano-Puche, J. (2021). Meta-framework of digital literacy: A comparative analysis of 21st-century skills frameworks. *Revista Latina De Comunicacion Social*, *79*, 76–109.
- Chen, C. M., Li, M. C., & Chen, Y. T. (2022). The effects of web-based inquiry learning mode with the support of collaborative digital reading annotation system on information literacy instruction. *Computers & Education*, *179*, 104428.
- Chetty, K., Qigui, L., Gcora, N., Josie, J., Wenwei, L., & Fang, C. (2018). Bridging the digital divide: Measuring digital literacy. *Economics*, *12*(1), 20180023.
- Clark, J., Falkner, W., Balaji Kuruvadi, S., Bruce, D., Zummo, W., & Yelamarthi, K. (2019). Development and implementation of real-time wireless sensor networks for data literacy education. In *Proceedings of the 2019 ASEE North Central Section Conference*, Morgan Town, WV, USA (pp. 22–23).
- Cote, T., & Milliner, B. (2016). Japanese university students' self-assessment and digital literacy test results. *CALL Communities and Culture–Short Papers from EUROCALL*, 125–131.
- Covello, S., & Lei, J. (2010). A review of digital literacy assessment instruments. *Syracuse University*, *1*, 31.
- Demirbag, M., & Bahcivan, E. (2021). Comprehensive exploration of digital literacy: Embedded with self-regulation and epistemological beliefs. *Journal of Science Education and Technology*, *30*(3), 448–459.
- Deutschschweizer Erziehungsdirektoren-Konferenz (D-EDK) (2014). *Lehrplan21 – Rahmeninformationen*. Luzern: D-EDK Geschäftsstelle.
- Dewi, C., Pahriah, P., & Purmadi, A. (2021). The urgency of digital literacy for generation Z students in chemistry learning. *International Journal of Emerging Technologies in Learning (IJET)*, *16*(11), 88–103.
- Da Silva, P. D., & Heaton, L. (2017). *Fostering digital and scientific literacy: Learning through practice*. First Monday.
- Divya, P., & Haneefa, M. (2018). Digital reading competency of students: A study in universities in Kerala. *DESIDOC Journal of Library & Information Technology*, *38*(2), 88–94.
- Ei, C. H., & Soon, C. (2021). *Towards a unified framework for digital literacy in Singapore*. IPS Work. Pap. 39.
- Elliott, K. C., & McKaughan, D. J. (2014). Nonepistemic values and the multiple goals of science. *Philosophy of Science*, *81*(1), 1–21.
- Erstad, O., Kjällander, S., & Järvelä, S. (2021). Facing the challenges of 'digital competence' a Nordic agenda for curriculum development for the 21st century. *Nordic Journal of Digital Literacy*, *16*(2), 77–87.

- Eshet-Alkalai, Y., & Soffer, O. (2012). Guest editorial—Navigating in the digital era: Digital literacy: Socio-cultural and educational aspects. *Educational Technology & Society*, 15(2), 1–2.
- Esteve-Mon, F., Llopis, M., & Adell-Segura, J. (2020). Digital competence and computational thinking of student teachers. *International Journal of Emerging Technologies in Learning (iJET)*, 15(2), 29–41.
- Ford, M. J. (2015). Educational implications of choosing “practice” to describe science in the next generation science standards. *Science Education*, 99(6), 1041–1048.
- Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). IEA international computer and information literacy study 2018 assessment framework (p. 74). Springer Nature.
- Gebhardt, E., Thomson, S., Ainley, J., & Hillman, K. (2019). Gender differences in computer and information literacy: An in-depth analysis of data from ICILS (p. 73). Springer nature.
- Gibson, P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany*, 105(12), 1953–1956.
- Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates’ evaluation of scientific information and arguments. *CBE—Life Sciences Education*, 11(4), 364–377.
- Ha, M., Park, H., Kim, Y. J., Kang, N. H., Oh, P. S., Kim, M. J., & Son, M. H. (2018). Developing and applying the questionnaire to measure science core competencies based on the 2015 revised national science curriculum. *Journal of the Korean Association for Science Education*, 38(4), 495–504.
- Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers’ and students’ gender role bias and students’ confidence in STEM fields. *Social Psychology of Education*, 20, 929–945.
- Holincheck, N., Galanti, T. M., & Trefil, J. (2022). Assessing the development of digital scientific literacy with a computational evidence-based reasoning tool. *Journal of Educational Computing Research*, 60(7), 1796–1817.
- Hug, B., & McNeill, K. L. (2008). Use of first-hand and second-hand data in science: Does data type influence classroom conversations? *International Journal of Science Education*, 30(13), 1725–1751.
- Jin, K. Y., Reichert, F., Cagasan, L. P., Jr., de La Torre, J., & Law, N. (2020). Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. *Computers & Education*, 157, 103968. <https://doi.org/10.1016/j.compedu.2020.103968>
- Kawasaki, J., & Sandoval, W. A. (2020). Examining teachers’ classroom strategies to understand their goals for student learning around the science practices in the Next Generation Science Standards. *Journal of Science Teacher Education*, 31(4), 384–400.
- Kerlin, C. K., McDonald, S. P., & Kelly, G. J. (2010). Complexity of Secondary Scientific Data Sources and Students’ Argumentative Discourse. *International Journal of Science Education*, 32(9), 1207–1225.
- Kim, H. S., Ahn, S. H., & Kim, C. M. (2019). A new IChT literacy test for elementary and middle school students in republic of Korea. *Te Asia-Pacific Education Researcher*, 28, 203–212.
- Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE—Life Sciences Education*, 18(2), es2.
- Kotzebue, L. V., Meier, M., Finger, A., Kremser, E., Huwer, J., Thoms, L. J., & Thyssen, C. (2021). The framework DiKoLAN (Digital competencies for teaching in science education) as basis for the self-assessment tool DiKoLAN-Grid. *Education Sciences*, 11(12), 775. <https://doi.org/10.3390/educsci11120775>
- Lazonder, A. W., Walraven, A., Gijlers, H., & Janssen, N. (2020). Longitudinal assessment of digital literacy in children: Findings from a large Dutch single-school study. *Computers & Education*, 143, 103681.
- Lichti, D., Mosley, P., & Callis-Duehl, K. (2021). Learning from the trees: Using project budburst to enhance data literacy and scientific writing skills in an introductory biology laboratory during remote learning. *Citizen Science: Theory and Practice*, 6(1), 1–12. <https://doi.org/10.5334/CSTP.432>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385.
- Mardiani, E., Mokodenseho, S., Matiala, T. F., Limbalo, S. S. A., & Mokodompit, N. Y. (2024). Implementation of Digital Science and Literacy Teaching in Developing Science Literacy in Middle School Students in Indonesia. *The Eastasouth Journal of Learning and Educations*, 2(01), 63–74.
- Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, 5(1), 67–90.

- Mason, L., Scrimin, S., Tornatora, M. C., Suitner, C., & Moè, A. (2018). Internet source evaluation: The role of implicit associations and psychophysiological self-regulation. *Computers & Education*, *119*, 59–75.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Ministry of Education (MOE), Ministry of Science and ICT (MSICT), & Korea Foundation for the Advancement of Science and Creativity (KOFAC). (2019). Scientific literacy for all Koreans: Korean science education standards for the next generation. Seoul: KOFAC.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517.
- National Research Council, Division of Behavioral, Board on Science Education, & National Committee on Science Education Standards. (1996). National science education standards. National Academies Press.
- National Research Council. (2003). *BIO 2010: Transforming undergraduate education for future research biologists*. National Academies Press.
- National Research Council. (2013). *Next Generation Science Standards: For States*. The National Academies Press.
- NGSS Lead States. (2013). Next generation science standards: For states, by states. National Academies Press.
- OECD. (2019). An OECD Learning Framework 2030. The Future of Education and Labor, 23–35.
- Oh, S. S., Kim, K. A., Kim, M., Oh, J., Chu, S. H., & Choi, J. (2021). Measurement of digital literacy among older adults: Systematic review. *Journal of Medical Internet Research*, *23*(2), e26145.
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196.
- Polizzi, G. (2020). Digital literacy and the national curriculum for England: Learning from how the experts engage with and evaluate online content. *Computers & Education*, *152*, 103859.
- Porat, E., Blau, I., & Barak, A. (2018). Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computers & Education*, *126*, 23–36.
- Qiao, C., Chen, Y., Guo, Q., & Yu, Y. (2024). Understanding science data literacy: A conceptual framework and assessment tool for college students majoring in STEM. *International Journal of STEM Education*, *11*(1), 1–21.
- Qin, J., & D'ignazio, J. (2010). The central role of metadata in a science data literacy course. *Journal of Library Metadata*, *10*(2–3), 188–204.
- Rodríguez-Becerra, J., Cáceres-Jensen, L., Diaz, T., Druker, S., Padilla, V. B., Perna, J., & Aksela, M. (2020). Developing technological pedagogical science knowledge through educational computational chemistry: A case study of pre-service chemistry teachers' perceptions. *Chemistry Education Research and Practice*, *21*(2), 638–654.
- Siddiq, F., Hatlevik, O. E., Olsen, R. V., Thronsdon, I., & Scherer, R. (2016). Taking a future perspective by learning from the past—A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, *19*, 58–84.
- Sobieraj, S., & Krämer, N. C. (2019). The impacts of gender and subject on experience of competence and autonomy in STEM. *Frontiers in Psychology*, *10*, 1432.
- Son, M., & Jeong, D. (2020). Exploring the direction of science inquiry education in knowledge-information based society. *School Science Journal*, *14*(3), 401–414.
- Son, M., Jeong, D., & Son, J. (2018). Analysis of middle school students' difficulties in science inquiry activity in view of knowledge and information processing competence. *Journal of the Korean Association for Science Education*, *38*(3), 441–449.
- Song, J., Kang, S. J., Kwak, Y., Kim, D., Kim, S., Na, J., & Jung, Y. J. (2019). Contents and features of “Korean Science Education Standards (KSES)” for the next generation. *Journal of the Korean Association for Science Education*, *39*(3), 465–478.
- Tsybulsky, D., & Sinai, E. (2022). IoT in project-based biology learning: Students' experiences and skill development. *Journal of Science Education and Technology*, *31*(4), 542–553.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, *52*(1), 234–246.

- Wolfe, E. W., & Smith, E. V., Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part I-instrument development tools. *Journal of Applied Measurement*, 8(1), 97–123.
- Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129, 41–54.
- Yasa, A. D., & Rahayu, S. (2023). A survey of elementary school students' digital literacy skills in science learning. In AIP Conference Proceedings (Vol. 2569, No. 1). AIP Publishing.
- Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, 36(1), 1–28.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Mihyun Son¹  · Minsu Ha² 

✉ Minsu Ha
msha101@snu.ac.kr

Mihyun Son
79algus@snu.ac.kr

¹ Future Innovation Institute, Seoul National University, 173 Seouldaehak-Ro, Siheung, Gyeonggi-Do, Republic of Korea

² Department of Science Education, Seoul National University, 1, Gwanak-Ro, Gwanak-Gu, Seoul, Republic of Korea