Understanding science data literacy: a conceptual framework and assessment tool for college students majoring in STEM

Cuilan Qiao[1], Yuqing Chen[1], Qing Guo[2*] and Yunwei Yu[1*]

Abstract

Background In the era defined by the fourth paradigm of science research, the burgeoning volume of science data poses a formidable challenge. The established data-related requisites within science literacy now fall short of addressing the evolving needs of researchers and STEM students. Consequently, the emergence of science data literacy becomes imperative. However, notwithstanding the escalating importance of science data literacy, a comprehensive definition and conceptual framework are still elusive, posing challenges to effective assessment.

Results This study analyzes the science literacy frameworks of six international organizations and countries, including the OECD, and reviews 16 data literacy-related literature sources identified using the PRISMA process. We also consider the characteristics of science data. Based on these sources, we clarify the connotation of science data literacy and construct a tailored conceptual framework for college students majoring in STEM. The framework undergoes two rounds of Delphi method refinement to enhance its applicability. Subsequently, we created and piloted a set of questions using physics, astronomy, geography, and other STEM subjects as examples to assess science data literacy. The revised assessment tool was then used in a formal test with 198 university students, employing Rasch modeling to evaluate its effectiveness. The tool's validity in assessing science data literacy was confirmed.

Conclusions This study offers a systematic and comprehensive conceptual framework for science data literacy tailored to STEM undergraduates. Endorsed by experts, the framework outlines essential literacies for STEM students in handling science data. The developed assessment tool enables educators to measure students' science data literacy levels and serves as a scientific guide to enhance their competencies in this area.

Keywords Science data literacy, Conceptual framework, Assessment tools, Delphi, Rasch

Introduction

Analyzing, understanding, and interpreting data have always been a crucial aspect of science literacy (Shaffer et al., 2019; Sholikah & Pertiwi, 2021). In the contemporary era characterized by the proliferation of big data, the sheer volume and diversity of data have surpassed the capabilities of manual analysis, and indeed, they have exceeded the capacities of traditional database management systems (Provost & Fawcett, 2013). Concomitant with the increasing power of computers, the ubiquity of networks, and the sophistication of computational and statistical methodologies for processing and analyzing large data sets, the field of data science has

*Correspondence:
Qing Guo
1287917960@qq.com
Yunwei Yu
yuyw@mail.ccnu.edu.cn
[1] College of Physical Science and Technology, Central China Normal University, Wuhan 430079, Hubei, China
[2] Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, Hubei, China

expanded significantly (Karpatne et al., 2017). In scientific research, the routine generation of vast amounts of data by contemporary devices worldwide is meticulously recorded, stored, and, more often than not, publicly disseminated. This development of data science within the context of big data has catalyzed a paradigmatic shift in the methodologies of scientific inquiry. Traditionally, scientific progress has been marked by the formulation of hypotheses or theories, followed by the collection of data to either corroborate or falsify these propositions (Karpatne et al., 2017). Nonetheless, in the age of big data, the continuous accumulation of data, often without pre-existing theoretical frameworks or specific hypotheses, presents novel opportunities for the discovery of new knowledge. Researchers can now harness large data sets to conduct simulations, engage in modeling, uncover previously unknown causal relationships, and articulate new theories (Karpatne et al., 2017; Kelling et al., 2009; Tansley & Tolle, 2009). Consequently, science research has progressed into the fourth paradigm era, propelled by this extensive science data (Tansley & Tolle, 2009). This shift necessitates enhanced data-related competencies as part of people's SL. To meet this evolving demand, the concept of science data literacy (SDL) has been introduced by several scholars (Qin & D'ignazio, 2010). This is particularly pertinent for college students majoring in STEM, as both their current academic pursuits and future professional careers are likely to involve direct engagement with massive volumes of science data. This study will use the SDL of college students majoring in physics, astronomy, and geography as an example of how to understand and improve the competencies of STEM college students in this critical area.

SDL, as defined by Qin and D'ignazio (2010), is "the ability to understand, use, and manage science data". Despite this definition, a universally recognized conceptual framework for SDL remains absent. This gap impedes a clear comprehension of the full scope of SDL. Therefore, developing a well-defined conceptual framework for SDL is vital for higher education institutions aiming to cultivate scientific and technological talent. Furthermore, the absence of a unified conceptual framework in the field of science education poses significant challenges in designing effective assessment tools. These tools are essential for evaluating students' proficiency in SDL and providing targeted instruction.

To address prevailing gaps, the imperative task is the systematic and scientific construction of a conceptual framework for SDL. In the fourth-paradigm era of science research, researchers are not solely engaged in designing experiments and collecting project-specific data; rather, their endeavors increasingly involve tapping into the extensive global repository of publicly available science data (Kelling et al., 2009; Tansley & Tolle, 2009). This approach aligns with practices in the big data era, where individuals navigate vast amounts of political, business, and other data types to address specific issues (Michener & Jones, 2012; Sander, 2020). Consequently, this study leverages the data literacy (DL) framework from the big data era to enhance the comprehension of SDL. Shields (2004), an early advocate for DL, defined it as the ability to access, manipulate, and present data. Since then, DL has gained increasing attention from researchers (Calzada & Marzal, 2013; Kippers et al., 2018; NAS, 2018), leading to a more developed conceptual framework (Pangrazio et al., 2019; Sander, 2020). However, it is crucial to recognize the distinction between general data and science data when constructing a framework for SDL. This study aims to integrate the data-related requirements in SL with the current understanding of DL to develop a comprehensive framework for SDL. Furthermore, the study aims to develop and assess an SDL evaluation tool that is customized for undergraduate students in STEM fields, including physics, astronomy, and geography. The research objectives are twofold: (1) to establish a conceptual framework for SDL specific to undergraduate students in STEM majors and (2) to create and validate an SDL assessment instrument for students in these majors, with physics, astronomy, and geography serving as illustrative disciplines.

Theoretical framework

Science data, science literacy, and science data literacy

Science data, comprising information collected and analyzed through experiments, observations, and calculations in science research (Demchenko et al., 2012; Fox & Hendler, 2011), include diverse examples like experimental data from physics, chemistry, and biology, observations of planetary movements, and atmospheric data. The primary distinction between science data and general data lies in their purpose: science data are often used to validate or refute scientific hypotheses, support or challenge theories, and uncover new knowledge (Fox & Hendler, 2011; Tenopir et al., 2011), while general data may have commercial, political, social, or other non-scientific applications (Fotopoulou, 2021; Katal et al., 2013). In addition, science data adhere to rigorous standards to ensure reliability and reproducibility (Wilkinson et al., 2016), unlike general data which may not always maintain such rigor, as evidenced by the use of ICT tools to capture various user behaviors in daily life (Yang et al., 2020). In the era of big data, the utilization of science data has become akin to that of general data. People typically assess data's value based on its utility in problem-solving, choosing the necessary data from a massive pool for this purpose (Pangrazio & Neil, 2019).

This approach to massive data usage has also permeated science research. With the development and use of large scientific instruments worldwide, research organizations and laboratories are generating significant amounts of science data continuously (Fataliyev & Mehdiyev, 2019). Consequently, this shift has altered the paradigm of science research: STEM professionals can now explore the extensive publicly available science data for data that can aid in solving their current problems or spark innovative research (Mustafee, et al., 2020; Tansley & Tolle, 2009).

It is widely acknowledged that the desired outcome of science education is SL (Siarovan et al., 2019). An individual who has SL is expected to possess a specific set of knowledge, skills, and attitudes, including the ability to explain phenomena scientifically, evaluate and design scientific inquiry, and interpret data and evidence scientifically (Council of the European Union, 2018; OECD, 2019; Sholikah & Pertiwi, 2021). The connotations of SL have led to the development of various conceptual frameworks, wherein the collection, analysis, interpretation, and argumentation of science data are recognized as pivotal elements (Shaffer et al., 2019; Sholikah & Pertiwi, 2021). However, existing SL requirements for data tend to emphasize skill acquisition, particularly students' ability to obtain and work with small amounts of data derived from inquiry experiments. With the exponential growth of science data, researchers are increasingly venturing beyond the laboratory to extract potential value from vast data sets for solving scientific problems of interest (Mustafee et al., 2020). Merely focusing on science data skills within the confines of the laboratory is no longer adequate to meet the demands of the evolving research paradigm driven by massive science data for future researchers. For instance, while they may possess the skills to work with small amounts of experimental data, they may face challenges when confronted with extensive data sets. This introduces a new set of challenges, such as identifying valuable science data for problem-solving, systematically organizing, managing, analyzing, and interpreting large-scale science data, and using publicly available data sets ethically. Hence, there is a pressing need to propose a systematic and well-developed conceptual framework for SDL that can effectively address these challenges.

Current research on SDL has yielded valuable insights. For instance, Qin and D'ignazio (2010) define it as "the ability to understand, use, and manage science data", although they do not propose a specific conceptual framework. Carlson et al. (2011) address SDL by constructing a framework based on a geoinformatics curriculum, emphasizing skills such as interpreting graphs and charts, drawing conclusions from data, and recognizing data misuse. However, this framework predominantly focuses on skill levels and lacks a systematic description of SDL. Another contribution comes from the science data lifecycle theory, which advocates for the comprehensive documentation and management of the entire lifecycle of science data from creation to disposal (Ball, 2012). Scientists benefit from this framework as it enables them to anticipate and plan actions required at each stage of the data application (Faundeen et al., 2014). The theory encompasses data planning, collection, management, analysis and visualization, sharing and preservation, discovery, and reuse (Michener & Jones, 2012; Qin & D'ignazio, 2010). While this theory informs the conceptual framework of SDL, it accentuates the skill dimension and lacks a fully developed connotation. In summary, existing research on SDL either lacks a fully developed conceptual framework or predominantly covers the skills dimension, resulting in an incomplete understanding of SDL. Consequently, there is an urgent need to construct a comprehensive SDL conceptual framework that transcends mere data skills.

Initial construction of a conceptual framework for science data literacy

The current conceptual frameworks of SL prescribe specific requirements for science data skills. When constructing conceptual frameworks for SDL, it is crucial to reference these existing frameworks to ensure alignment with established scientific principles. This study conducts an analysis of the conceptual underpinnings of SL within prominent international organizations and representative countries. This analysis includes the OECD's (2019) PISA 2018 framework for assessing SL, the Council of the European Union's (2018) key competencies for lifelong learning concerning SL, Schneegans et al. (2021) in UNESCO report on SL, the National Academies of Sciences, Engineering, and Medicine's (2016) co-authored work "Science Literacy: Concepts, Contexts, and Consequences", the General Office of the State Council of the PRC (2021), and the Government of Canada's (2021) perspectives on SL. Relevant elements extracted from these sources are summarized in Table 1. The analysis reveals that the data-related requirements for SL primarily emphasize the skill level in using data, with only a few conceptual frameworks addressing data ethics. As discussed in the preceding subsection, these elements prove insufficient to meet the challenges of addressing scientific problems within vast amounts of science data. Therefore, there is a need to incorporate DL from the era of big data to refine conceptual frameworks for SDL.

The term "data" in DL encompasses general data, including various forms of business and political information (Pangrazio & Selwyn, 2019; Fotopoulou, 2021). Much like SDL, both are responses to the

Table 1 Foundations of the conceptual framework for SDL

	Data awareness			Data skills											Ethics
	Sensitive	Value	Security	Access	Storage	Assessment	Organization	Management	Processing	Visualization	Analysis	Explanation	Presenting	Share	Morality
Science literacy (SL)															
OECD (2019)				✓		✓			✓		✓	✓			
Schneegans (2021)				✓		✓					✓	✓			
Council of the European Union(2018)						✓			✓		✓	✓			✓
National Academies of Sciences, Engineering, and Medicine (2016)						✓			✓		✓	✓			✓
General Office of the State Council of PRC (2021)				✓		✓			✓		✓	✓	✓	✓	✓
Government of Canada (2021)						✓					✓	✓			
Data literacy (DL)															
Shields (2004)				✓	✓	✓			✓				✓		
Vahey, et al. (2006)						✓			✓		✓				
Australian Bureau of Statistics (ABS, 2010)	✓			✓		✓	✓		✓	✓	✓	✓	✓	✓	
Carlson et al. (2011)			✓	✓	✓		✓	✓	✓	✓	✓	✓			✓
Calzada and Marzal (2013)		✓		✓	✓	✓		✓	✓	✓	✓	✓	✓		✓
D'Ignazio and Bhargava (2015)	✓			✓			✓	✓	✓		✓	✓		✓	✓
Gummer et al. (2015)				✓				✓	✓		✓	✓			
Wolff et al. (2016)	✓	✓	✓	✓				✓	✓		✓	✓	✓	✓	✓

Table 1 (continued)

| | Data awareness | | | Data skills | | | | | | | | | | | Ethics |
	Sensitive	Value	Security	Access	Storage	Assessment	Organization	Management	Processing	Visualization	Analysis	Explanation	Presenting	Share	Morality
Koltay (2017)		✓		✓	✓	✓		✓	✓		✓	✓		✓	✓
Gibson et al. (2018)	✓	✓		✓		✓		✓	✓	✓	✓	✓		✓	
Pangrazio and Selwyn (2019)	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓		✓	✓
Sander (2020)	✓	✓	✓	✓					✓			✓		✓	✓
Bowler et al. (2020)		✓		✓	✓	✓	✓	✓	✓		✓			✓	✓
Carmi et al. (2020)	✓	✓	✓	✓		✓		✓	✓			✓		✓	✓
Fotopoulou (2021)	✓	✓		✓		✓			✓	✓	✓	✓	✓	✓	✓
Gebre (2022)	✓	✓	✓	✓					✓	✓	✓	✓		✓	✓

challenges posed by big data. The process by which individuals utilize general data to accomplish tasks mirrors the approach researchers take when utilizing massive amounts of science data to address scientific problems. Consequently, DL proves valuable in constructing a conceptual framework for SDL. To identify pertinent literature, this study employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) process (Page et al., 2021) as a guide. The Web of Science core database was selected for the search to ensure the authority of the literature. Using the search term "'data literacy' OR 'data literacies'" within the date range "1985-01-01 to 2022-12-31", a total of 379 related literatures were retrieved. The study then screened English literature focusing on DL for students or researchers and containing a clear conceptual framework. Through examination of titles and abstracts, 16 pieces of literature were identified as eligible, one of which could not be accessed in the original language. After reading the full text of the remaining 15, five articles that did not meet screening requirements were excluded. Subsequently, six relevant works were supplemented by other means, such as citations of included articles. Ultimately, 16 high-quality and representative papers were included in the analysis. These literatures were organized chronologically, and their proposed conceptual framework for DL was dissected into core components for analysis, as detailed in Table 1. A thorough review of each piece of literature was conducted to identify the representations defining and structuring the conceptual framework for DL. Keywords used in these conceptual framework representations were captured and summarized. To ensure consistency, multiple levels of keyword consolidation were performed, grouping elements with similar connotations. For instance, terms such as "collect data", "acquire data", and "access data" were consolidated into a single element labeled "access", which, along with other elements, can be considered part of the data skills dimension.

Evidently, DL initially centered on fundamental data skills but has undergone evolution, with growing attention from researchers toward data ethics and data awareness. After the integration of keywords in Table 1, the ideas of data sensitivity, recognizing data value, and protecting data security can be collectively referred to as data-related awareness. This yields the first dimension of SDL, i.e., science data awareness, which is defined as human perception and understanding of science data, and includes three elements: data acuity awareness (DAA), data value awareness (DVA), and data security awareness (DSA).

Data-related skills are integral to both SL and DL, representing a core dimension within the construct of SDL.

This dimension is characterized by the proficiency to utilize science data in the pursuit of scientific inquiry and its practical application. Through a thorough literature review and synthesis, we have determined that the ability to access data includes the processes of identifying and gathering data. In addition, the organization and management of data involve the assessment and strategic storage of data. These activities are often closely linked, and thus, they are consolidated into a single category: data organization and management skills. Data processing and analysis are frequently interwoven, leading to their combination into a unified skill set: data processing and analysis skills. Presentation of data results is often a process of sharing and communication, so we combine presentation and sharing into data communication and analysis skills. Moreover, the visualization of data and the interpretation of data outcomes are essential for researchers engaging with science data. In summary, the dimension of science data skills is comprised of six distinct elements: data finding and collection skills (DFACS), data organization and management skills (DOAMS), data processing and analysis skills (DPAAS), data visualization skills (DVS), data interpretation skills (DIS), and data communication and sharing skills (DCASS).

The analysis of literature highlights the critical importance of the ethical use of data for researchers. As big data continues to expand, data-related laws and regulations are being refined to set appropriate boundaries for data use. Hence, it is imperative for data users not only to adhere to ethical guidelines for science data usage but also to comply with legal and regulatory requirements for data access. In response to these findings, this study proposes incorporating legal and regulatory perspectives to complement the ethical perspectives that emerged from the literature review. Thus, the dimension of science data regulations and ethics is conceptualized to encompass two distinct elements: data laws and regulations (DLAR) and data ethics (DE). This dimension is defined by adherence to the legal, regulatory, and ethical protocols that govern the collection, use, and dissemination of science data.

In summary, SDL has been initially categorized into three dimensions: science data awareness, science data skills, and science data regulations and ethics, including a total of 11 elements, as shown in Fig. 1, which creates a multidimensional and comprehensive framework for SDL. The content analysis shows that when contemplating SL alongside DL in the era of big data in a holistic manner, SDL can be delineated as a composite of individuals' awareness, capacity for application, and adherence to ethical norms in addressing scientific problems rationally utilizing massive science data.

Science Data Literacy		
Science Data Awareness	**Science Data Skills**	**Science Data Regulations & Ethics**
• Data acuity awareness • Data value awareness • Data security awareness	• Data finding & collection skill • Data organization & management skill • Data processing & analysis skill • Data interpretation skill • Data deduction skill • Data communication & sharing skill	• Data laws & regulations • Data ethics

Fig. 1 Conceptual framework of SDL (preliminary)

Methods

Procedure

A preliminary version of the conceptual framework for SDL presented in this study will be revised and refined. The Delphi method will be utilized to gather expert consensus. The Delphi method involves multiple rounds of soliciting opinions from experts, collecting, analyzing, and revising feedback to reach conclusions with more unanimous opinions (Skulmoski et al., 2007). For this study, experts were engaged in multiple rounds of Delphi surveys, receiving e-mail questionnaires about their agreement on the dimensions and elements of SDL. Experts were asked to assess the appropriateness and importance of each dimension and element and to provide suggestions for revisions. Each survey round allowed 2–3 weeks for responses, and the feedback guided the revisions of the conceptual framework until expert agreement was achieved in a particular round.

With the establishment of a broadly accepted conceptual framework, this study will use it as a foundation to develop a testing tool for the level of SDL applicable to college students in related majors, using physics, astronomy, and geographical sciences as examples. Initially, 3–4 multiple-choice questions were created for each element, and experts collaboratively refined the test instrument through joint discussions and redaction. Subsequently, a group of college students participated in the test, and the Rasch model was used to determine whether test questions could measure students' SDL. The Rasch model is a widely recognized method for evaluating the quality of measurement instruments, providing insights into the accuracy of each question in reflecting the subject's competence (Weller, 2013). Based on the Rasch analysis outcomes, adjustments were made to the questions. Finally, another group of college students completed the modified test and applied the Rasch model again to determine the validity of the modified question.

Materials

The Delphi questionnaire consists of two versions: one for the initial round and another for each subsequent round. Each version includes the following sections: a greeting, a survey for basic information (Fig. 2a), an introduction to the connotation of the three dimensions in SDL (Fig. 2b), scoring tables for the three dimensions and their elements (Fig. 2c,), and a section for expert explanations and recommendations (Fig. 2e). The rating form uses Likert's five-point scoring method ("1" indicates very unimportant, and "5" indicates extremely important). If an expert scores less than 2 points for a dimension or element, they must provide a written explanation in the designated section. Experts can also offer suggestions for modifying dimensions or elements in the suggestion section. In subsequent rounds, the questionnaires were revised based on expert opinions from the previous round. The new round of questionnaires includes all elements of the new conceptual framework for SDL, along with the revised responses to the previous round of expert opinions (Fig. 2f). Experts then determine new scores based on these revisions. The response to revisions consists of two parts: a description of the main changes to the framework and a point-by-point response to expert suggestions.

Two versions of the SDL assessment test questions were used. The initial version comprised 33 multiple-choice questions, which were filtered down to 22 questions for the official version based on pilot test results. Each element of SDL was assessed by two items; for
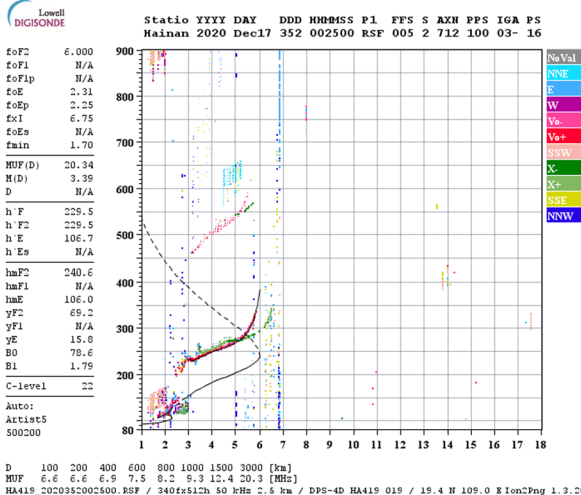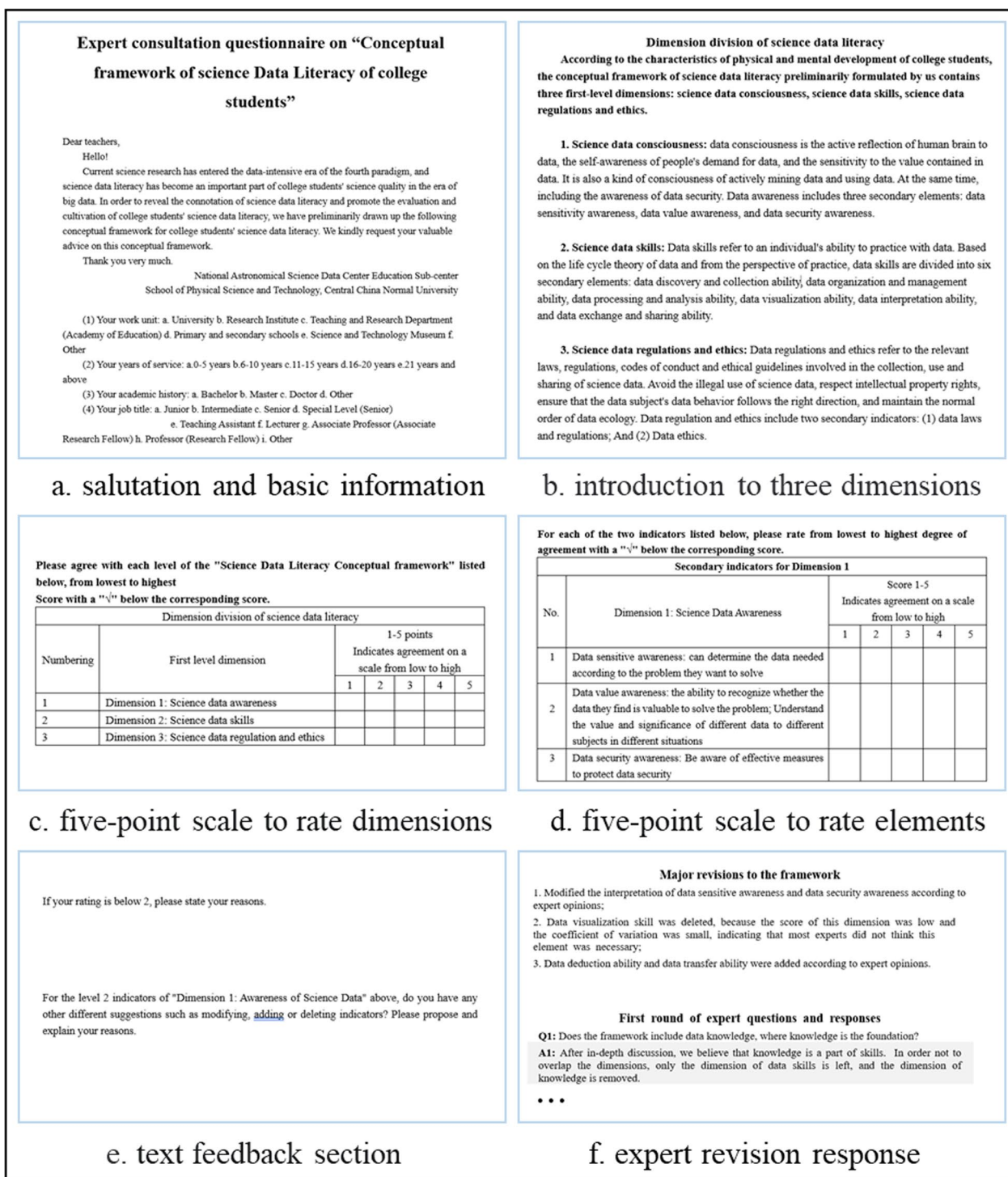
Fig. 2 Structure of the Delphi questionnaire

example, items 1 and 2 assessed a student's DAA. The multiple-choice format presented students with a question stem and four options, requiring them to select the most appropriate answer. Sample questions from the official version can be found in Table 2. It is worth noting that having SDL is necessary for every college student in STEM-related disciplines. However, science data are distinct in different subject areas, and students in each

Table 2 Sample test questions for SDL assessment

Dimension	Sample item
Science data awareness	Hong is a graduate student who collects research data from her lab. Which of the following behaviors is best for protecting data security? () A. Copy data with your own portable hard drive B. Data is stored in the local database of the laboratory C. Send the data to other students' personal computers D. Upload the required data from the database to a third-party web disk
Science data skills	Ming has made a data image, as shown in the figure below. He wants to introduce the meaning of the image to others, which of the following is not required? () A. Describes the meanings of the horizontal and vertical axes B. This section describes each parameter in the list C. Describes the software that processes the data D. Introduce the meaning of each color
Science data regulation and ethics	Hong wants to use the data in a data center, but she does not have the download permission, which of the following is correct? () A. Find someone online to download it for you B. Send an email to the relevant department to apply for permission C. Make up your own data based on previous data D. Look for unofficial channels to buy data online

The graphs for the examples in the table are from the Meridian Project Data Center (2023)

discipline face unique contexts when engaging with and learning about science data. In constructing the assessment questions, we paid particular attention to the use of authentic science data and contexts to ensure that the test questions realistically reflected actual research scenarios. Given that the student participants in this study, as well as the team that developed the assessment tool, were primarily from disciplinary fields such as physics, astronomy, and geography, we designed some of the assessment test questions using real science data and scientific contexts from these disciplines. This design not only makes the test questions more relevant and specialized, but also helps students better understand and adapt to the science data in their subject areas. It should be emphasized that the core purpose of these test questions is to be used to assess students' ability to understand and apply science data, and the data and contextual information involved are intended to simulate the scenarios that researchers face when dealing with real science data, to help students better understand the realities of scientific research.

Participants

To ensure the final conceptual framework for SDL meets the requirements of scientific work and college students' educational needs, this study sought input from a diverse group of experts. A total of 33 experts from various fields participated in the survey, including college science education specialists (professors and associate professors engaged in the education of science subjects such as physics, chemistry, biology, astronomy, geography, etc. at universities), university STEM teachers, researchers from research institutes (e.g., researcher at the Institute of Science Education), data scientists (e.g., professors in the field of physical data science), and science and

technology museum staff. This expert team represents a broad spectrum of expertise and authority in the science field. Approximately 45% of the experts had more than 5 years of experience, with 7 experts having over 20 years of experience. These experts possess knowledge of data science and are involved in teaching courses or conducting experiments related to science research for college students majoring in STEM. They are well aware of students' utilization of science data. In addition, experts from various fields can contemplate the type of SDL that STEM majors should possess from different perspectives. Synthesizing the opinions of experts across these fields can render the conceptual framework of the SDL we constructed more comprehensive.

Since the developers of the test questions in this study were mainly from the subject areas of physics, astronomy, and geography, a group of college students from related majors was also selected for this study to verify the validity and reliability of this SDL testing tool. 83 physics major students from a university in central China voluntarily participated in the first round of test question testing. The subsequent official test involved 198 students from the Astronomy Association of a university in central China, more than 90% of whom majored in physics, astronomy, and other related majors. Among the participants, 103 were male, and 95 were female, spanning various undergraduate levels from the first to the fourth year, as well as graduate students. Approximately 2.0% were freshmen, 78.3% were sophomores, 18.7% were juniors, and 1.0% were seniors and graduate students.

Analysis

The data analysis involved two phases, aligned with the research objectives. The first phase aimed to identify the elements of SDL through expert consultation. Elements with a mean score of 4 or higher on Likert's five-point scale were considered "very important" by 80% of the experts (Langlands et al., 2008; Law & Morrison, 2014). In addition, high expert agreement was determined using the criteria of interquartile coefficient of variation (CQV) ≤ 0.2 (Quyên, 2014) and standard deviation (SD) < 1 (Liao et al., 2017). Therefore, dimensions and elements with mean scores greater than or equal to 4 and CQV less than or equal to 0.2 and SD less than 1 were retained.

The second stage aims to assess the quality of the SDL assessment test questions. The official version of the quiz, revised based on the pilot test, will be presented in the main study. The official test questions were analyzed using the Rasch model analysis software Winsteps 5.3.0.0, which allows the most detailed assessment of the quality of the project (Linacre, 2019; Pedaste et al., 2023). The holistic analysis of the test questions involved several

assessment criteria. (1) Unidimensionality: Good unidimensionality requires that the unexplained variance in the first comparison be less than 3 and account for less than 15% of the variance (Linacre, 2019). In addition, the ratio of the variance explained by the test to the percentage of variance accounted for by the unexplained first eigenvalue should be greater than 3 (Hays et al., 2000). (2) Overall reliability and separation: generally, item reliabilities greater than 0.9 and separations greater than 4.0 were required (Malec et al., 2007). (3) Test information curve: the peak value of the test information curve should be greater than 5, corresponding to classical test theoretical reliability greater than 0.8 (Young et al., 2013). The item-level assessment focused on the following criteria: (1) item difficulty distribution, with a mean item difficulty set to 0. (2) Fit index (Infit and Outfit) reflecting the degree of fit of the items to Rasch's theoretical model, with an ideal MNSQ fit range between 0.5 and 1.5 (O'Connor et al., 2016). (3) Point-measurement correlation (PT-Measure CORR.), which indicates the correlation between the model-estimated item scores and the actual values. An acceptable correlation was considered greater than 0.2, while an ideal correlation was greater than 0.3 (Pedaste et al., 2023). In addition, this study developed a Wright Map to analyze the correspondence between subjects' abilities and item difficulty. To further explore whether these items have the same effect on subjects of different genders, DIF was used as a test. Items with an absolute value of DIF contrast greater than 0.5 can be determined to function differently in gender (Fox & Bond, 2015).

Results

Conceptual framework of science data literacy

Analysis of the results of the expert consultation

The first round of the Delphi survey involved sending 33 questionnaires, of which 30 were recovered. After excluding one questionnaire with missing and outlier values, 29 questionnaires were effectively recovered. The results of this round are summarized in Table 3. The mean scores for all dimensions were greater than 4, with CQV values below 0.2 and SD values less than 1, indicating a consensus among experts regarding the importance of these dimensions. As a result, all three dimensions were retained. However, in the case of data visualization competence, despite the mean scores being greater than 4 and CQV values being less than 0.2 for all elements, experts expressed a general agreement that it should be included within data processing and analysis competence. Therefore, data visualization competence was removed from the second round of the questionnaire. Moreover, experts proposed modifications to the specific interpretations of data sensitivity awareness, data security awareness, data

Table 3 Results of the expert consultation

Dimension	Round 1 Mean (CQV/SD)	Round 2 Mean (CQV/SD)	Element	Round 1 Mean (CQV/SD)	Round 2 Mean (CQV/SD)
Science data awareness	4.69 (0.03/0.53)	4.87 (0.00/0.34)	DAA	4.41 (0.11/0.84)	4.74 (0.00/0.54)
			DVA	4.72 (0.00/0.71)	4.78 (0.00/0.42)
			DSA	4.48 (0.11/0.88)	4.64 (0.08/0.66)
Science data skills	4.69 (0.00/0.59)	4.91 (0.00/0.29)	DFACS	4.72 (0.00/0.65)	4.87 (0.00/0.34)
			DOAMS	4.72 (0.00/0.60)	4.91 (0.00/0.29)
			DPAAS	4.79 (0.00/0.48)	4.78 (0.00/0.52)
			DVS	4.59 (0.11/0.69)	Deleted
			DIS	4.76 (0.00/0.59)	4.78 (0.00/0.52)
			DDS	None	4.57 (0.11/0.66)
			DCASS	4.62 (0.11/0.69)	4.74 (0.00/0.54)
Science data regulation and ethics	4.55 (0.11/0.74)	4.78 (0.00/0.52)	DLAR	4.86 (0.00/0.31)	4.91 (0.00/0.29)
			DE	4.79 (0.00/0.48)	4.87 (0.00/0.34)

organization and management skills, and data interpretation skills. In addition, experts suggested the inclusion of data deduction skills (DDS) within the science data skills dimension. Subsequently, the second round questionnaire was modified according to the experts' comments, and detailed itemized responses were provided in response to the experts' suggestions.

The second round of the Delphi survey involved experts who had effectively completed the first round of questionnaires. Out of 29 distributed questionnaires, 23 were recovered and analyzed, with the results presented in Table 3. The mean scores for all dimensions and elements were greater than 4, while the CQV value was below 0.2 and the SD value was less than 1. These findings indicate a high level of consensus among the experts regarding the entire conceptual framework. Comparing the second round to the first round, it is evident that, except for the newly added "data deduction skill (DDS)" and the deleted "data visualization skill (DVS)", the average scores for all dimensions and elements were higher. This suggests that the revised conceptual framework is perceived by experts to be more scientifically robust and that the interpretation of the elements is clearer and more reasonable. In conclusion, the second round of the Delphi survey supports the notion that the revised conceptual framework is more scientifically sound, provides clearer explanations for the elements, and strengthens the importance of the dimensions and elements compared to the initial version.

Finalized conceptual framework

Following a comprehensive literature review, this study established a concise and logical conceptual framework for SDL in college students majoring in STEM. The framework was refined through two rounds of Delphi surveys, ultimately achieving expert consensus. It consists of three dimensions and 11 elements, of which science data awareness relates to human perception and understanding of science data, science data skills include the competencies required to use science data for scientific inquiry and practice, and science data regulations and ethics relate to the legal, regulatory, and ethical norms that govern the collection, use, and sharing of science data. The categorization of the eleven elements and their specific explanations are shown in Table 4.

Assessment tool for science data literacy

The results presented in this section showcase the official version (22-question version) of the Rasch test. A brief description of the 22 questions on the official version of the SDL's assessment test can be found in the Appendix.

Unidimensionality

The assessment of unidimensionality using PCAR yielded satisfactory results, as evidenced by the Unexplained variance in the 1st contrast of this test series being 2.1, with a ratio of 7%. These values meet the criteria of being less than 3 and below 15%, respectively. In addition, the ratio of the variance shared between the explained and unexplained first eigenvalues is ~ 3.6, satisfying the requirement of being greater than 3. These tests provide compelling evidence for the unidimensionality of the assessment.

Overall reliability, separation, and test information curves

Regarding the overall reliability of the SDL assessment items, Cronbach's alpha coefficient was calculated to be 0.95, surpassing the recommended threshold of 0.9. Moreover, the item separation reached 4.28, exceeding

Table 4 Conceptual framework for SDL (completed version)

| Dimension	Element	Explicit explanation
Science data awareness	DAA	Acutely aware of what kind of science data is needed for problem-solving when presented with a science problem
	DVA	Recognize the value and significance of different science data for addressing a particular science question
	DSA	Understand the risks of science data leakage and implement protective measures
Science data skills	DFACS	Retrieve, collect, evaluate, and select relevant science data from various sources
	DOAMS	Effectively store, arrange, organize, and classify science data to improve utilization efficiency
	DPAAS	Utilize common methods and tools for data processing, and perform analysis, modeling, and visualization
	DIS	Interpret visualization images to understand the information conveyed by science data and the significance of analysis results
	DDS	Utilize science data and processed conclusions as a basis for decision-making
	DCASS	Develop skills in communicating, expressing, and sharing science data, and effectively convey the value of data analysis work
Science data regulation and ethics	DLAR	Understand codes of conduct and ethics for science data use, ensuring proper data behavior and maintaining data ecology
	DE	Comply with relevant laws and regulations for science data collection, use, and sharing, respecting intellectual property rights and preventing illegal data use
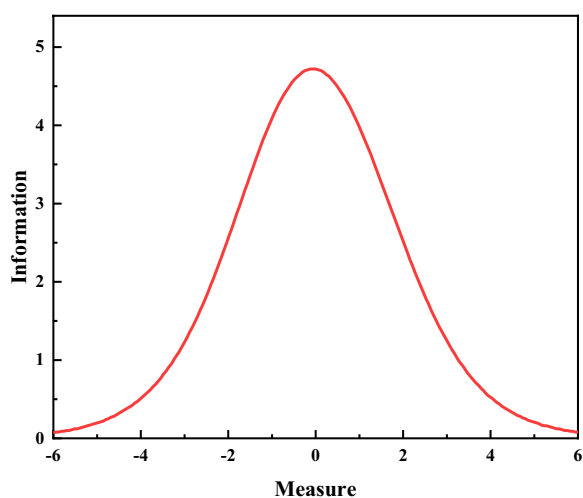
Fig. 3 Test information curve

the desirable value of 4. Figure 3 depicts the test information curve of the assessment questions, illustrating a peak value of around 5, which corresponds to a Cronbach's alpha value of ~ 0.8. In conclusion, these test results unequivocally demonstrate the high reliability of the revised official test questions.

Item difficulty, fit index, point-measurement correlation

Table 5 presents essential information on individual item fitting, including item serial number, total item score, number of participants, item difficulty, standard error of difficulty, Infit MNSQ, Outfit MNSQ, and PT CORR. The items are ordered in descending order of difficulty (measure value) ranging from 1.51 to − 1.28. Item 5 is the most difficult, while item 13 is the least difficult. The standard error of difficulty for all items falls within the range of 0.16 to 0.24.

In terms of fit indices, all items exhibit Infit MNSQ values ranging from 0.70 to 1.20 and Outfit MNSQ values ranging from 0.43 to 1.39. These values are within the required range of 0.5–1.5, indicating that the items generally meet the fitting requirements. Only Item 19 has an Outfit MNSQ value of 0.43, slightly below 0.5. Nevertheless, this value is close to the desired threshold, and Item 19's Infit MNSQ value is within an acceptable range (0.69). Moreover, subsequent point measurements related to the indexes also fall within acceptable limits (0.65). Therefore, Item 19 can also be considered acceptable. Furthermore, all items demonstrate point measure correlation values greater than 0.2, with the vast majority of items (90.9%) having point measure correlation values exceeding 0.3. This evidence supports the conclusion that all items in the revised formal test are deemed acceptable and high quality.

Wright map

The wright map of quiz results is depicted in Fig. 4 The plot utilizes a logarithmic scale on the center axis, specifically the logit scale. Participants ($n = 198$) are shown on the left side of the graph, ranked in descending order based on their ability values. On the right side, the 22 items are also arranged in descending order based on

Table 5 Item fit index

Item	Total score	Count	Measure	Mode S.E	Infit MNSQ	Outfit MNSQ	PT CORR
5	83	198	1.51	0.16	1.23	1.52	0.22
6	88	198	1.39	0.16	1.13	1.14	0.33
18	93	198	1.26	0.16	0.91	0.86	0.50
11	99	198	1.11	0.16	1.04	1.00	0.40
4	114	198	0.73	0.16	1.11	1.19	0.34
1	116	198	0.68	0.16	1.09	1.17	0.35
12	122	198	0.53	0.16	1.01	0.99	0.42
16	127	198	0.39	0.16	1.13	1.19	0.33
9	132	198	0.25	0.17	1.07	1.07	0.37
8	142	198	− 0.04	0.18	1.19	1.22	0.28
14	143	198	− 0.07	0.18	0.91	0.87	0.48
10	147	198	− 0.20	0.18	1.03	0.97	0.40
15	150	198	− 0.30	0.18	0.90	0.79	0.49
21	152	198	− 0.37	0.19	0.83	0.73	0.53
2	154	198	− 0.44	0.19	0.91	0.90	0.46
17	157	198	− 0.55	0.19	0.85	0.81	0.50
3	162	198	− 0.75	0.20	1.05	1.03	0.36
22	163	198	− 0.79	0.21	0.90	0.75	0.47
7	166	198	− 0.92	0.21	1.11	1.15	0.31
20	166	198	− 0.92	0.21	0.78	0.60	0.55
19	172	198	− 1.22	0.23	0.69	0.43	0.59
13	173	198	− 1.28	0.24	0.89	0.66	0.46

their difficulty. The figure indicates a relatively even distribution of item difficulty. Overall, the distribution of participants and item difficulty appears to be reasonably balanced, although the latter slightly skews lower. This suggests that the tested items align well with the cognitive level of college students, though there is room to marginally increase the difficulty further.

DIF test
Table 6 presents the results of the DIF test, where gender is used as a variable. It reveals a gender-based contrast in responses, particularly noticeable in Item 2, Item 4, Item 5, Item 14, Item 17, Item 19, and Item 20, as indicated by their DIF Contrast values exceeding the absolute threshold of 0.5. Specifically, Item2, belonging to the DAA element, along with Item14 (DIS element), Item17 (DCASS element), and Items 19 and 20 (both in DLAR element), all show negative DIF Contrast values greater than 0.5, indicating a gender-related difference in performance. The superior performance of female students on these items suggests a propensity to excel in recognizing pertinent data, as well as in interpreting, communicating, and sharing the data. Conversely, Item 4 (DVA element) and Item 5 (DSA element) also yielded DIF Contrast values

exceeding 0.5, indicating that male students may have a comparative advantage in these dimensions, implying a stronger aptitude for recognizing the value of data and upholding robust data security practices.

Discussion
SDL has emerged as a pivotal aspect of science research and innovation, holding significant implications for both scientists and students pursuing STEM-related careers. Despite its growing importance, there currently exists no comprehensive conceptual framework for SDL, and corresponding assessment tools are notably lacking. This study aims to address these gaps by constructing a conceptual framework of SDL tailored for college students majoring in STEM, encompassing three dimensions and eleven elements. In addition, a suite of SDL assessment tools was developed based on this framework. The comprehensiveness and validity of both the conceptual framework and the assessment tool were validated through the application of the Delphi method and Rasch model. The outcomes of this study contribute substantially to the assessment and cultivation of SDL among college students, particularly those specializing in STEM disciplines.
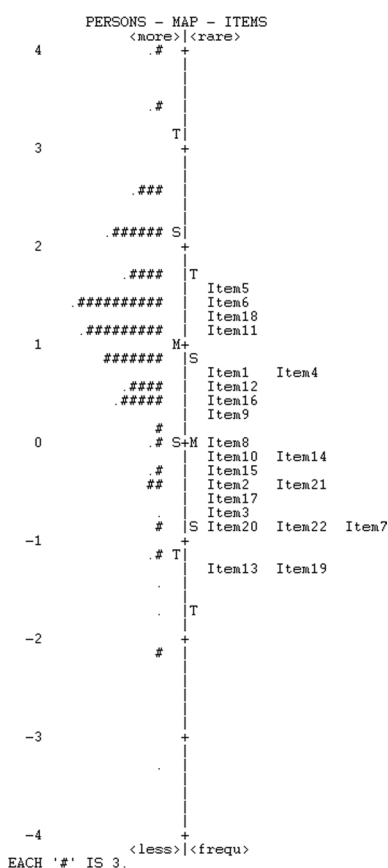
Fig. 4 Wright map

Importance and conceptual framework of science data literacy

SDL encompasses the skills required for engaging with data within the realm of SL, encompassing tasks such as assessment, analysis, and comprehension of data (OECD, 2019; Sholikah & Pertiwi, 2021). Furthermore, it involves elements of DL, including awareness of data, data collection, organization, management, and the capacity to analyze arguments based on data (Calzada & Marzal, 2013; Gebre, 2022; Wolff et al., 2016). The characteristics of science data, such as variety, volume, and specialization, are also considered within this framework (Wilkinson et al., 2016). Recognizing the centrality of data in science research and practice (Ball, 2012; Michener & Jones, 2012), students equipped with SDL possess the ability to discern the value of extensive science data. They are adept at comprehending and analyzing data to derive accurate and reliable conclusions, thereby enhancing the quality and efficiency of research and practice (Faundeen et al., 2014). Moreover, SDL empowers students to better understand and expound upon scientific phenomena and problems, thereby contributing to the advancement and progress of science and technology (Siarova et al., 2019).

In developing a conceptual framework for SDL, expert feedback played a crucial role. It affirmed the inclusion of all three dimensions: science data awareness, science data skills, and science data regulations and ethics. This trifold framework is pivotal for STEM undergraduates, mirroring the practices of scientists who extensively use science data in their research. Firstly, scientists must possess an awareness of data's value and security, understand the types of data needed for research, and evaluate the data's relevance to problem-solving (Kjelvik et al., 2019). Second, they require proficiency in collecting, organizing, managing, and analyzing science data to derive meaningful insights (Rüegg et al., 2014). In addition, adherence to legal and ethical standards, including the maintenance of intellectual property rights, is imperative when handling science data (Wilkinson et al., 2016). Notably, during the expert consultation process, the introduction of "data deduction skills" into the science data skills dimension was proposed. This skill encompasses the ability to interpret data and make informed conclusions, crucial for decision-making. Experts stressed the significance of this competency for aspiring scientists and STEM practitioners, aligning with the current emphasis on data-driven decision-making in the era of big data (Trantham et al., 2021; Wolff et al., 2016).

The conceptualization and framework of SDL developed in this study build upon and extend existing studies. Prior studies into SDL have largely centered on the competencies associated with the use of science data. For instance, Qin and D'ignazio's (2010) research concentrated on the abilities to comprehend, apply, and manage science data, while the science lifecycle theory (Michener & Jones, 2012) accentuates skills in the acquisition, curation, and analysis of such data. In contrast, the definition of SDL and the conceptual framework proposed in this study further expands the connotation of SDL while inheriting the previous emphasis on science data skills. By integrating insights from SL, DL, and the unique attributes of science data, this study not only deepens the understanding of science data skills, but also adds two new dimensions of science data awareness and science data regulation and ethics. This expanded framework shifts the focus of SDL from a sole concentration on skills to a more inclusive perspective that encompasses awareness, skills, regulations, and ethics, thereby promoting a more comprehensive appreciation of SDL.

It is noteworthy that while the conceptual framework of SDL developed in this study shares some elements with the conceptual framework of DL, they are fundamentally distinct. Firstly, they center on different types of data. SDL is oriented towards science data rather than general data such as those related to business and politics. Second, they cater to different audiences, with DL positioned

Table 6 Gender DIF

Element	Item	Gender	DIF Measure	Gender	DIF Measure	DIF Contrast
DAA	1	Female	0.80	Male	0.55	0.25
	2	Female	− 0.95	Male	− 0.08	**− 0.87**
DVA	3	Female	− 0.84	Male	− 0.68	− 0.16
	4	Female	1.22	Male	0.16	**1.06**
DSA	5	Female	1.84	Male	1.14	**0.70**
	6	Female	1.31	Male	1.47	− 0.16
DFACS	7	Female	− 0.75	Male	− 1.07	0.32
	8	Female	− 0.13	Male	0.04	− 0.17
DOAMS	9	Female	0.45	Male	0.04	0.41
	10	Female	− 0.06	Male	− 0.33	0.27
DPAAS	11	Female	1.17	Male	1.04	0.13
	12	Female	0.65	Male	0.39	0.26
DIS	13	Female	− 1.47	Male	− 1.16	− 0.31
	14	Female	− 0.57	Male	0.33	**− 0.90**
DDS	15	Female	− 0.26	Male	− 0.33	0.07
	16	Female	0.50	Male	0.27	0.23
DCASS	17	Female	− 0.84	Male	− 0.33	**− 0.51**
	18	Female	1.31	Male	1.20	0.11
DLAR	19	Female	− 2.07	Male	− 0.83	**− 1.24**
	20	Female	− 1.64	Male	− 0.54	**− 1.10**
DE	21	Female	− 0.49	Male	− 0.27	− 0.22
	22	Female	− 1.06	Male	− 0.61	− 0.45

Bold value indicates that there is a significant gender difference in the question

as a general literacy applicable to all citizens (Fotopoulou, 2021), while SDL is tailored for current or prospective scientists, including researchers, engineers, and STEM students. Finally, although these two frameworks feature similar dimensions, they diverge in their specific connotations. For example, DL emphasizes understanding and decision-making based on data (Carey et al., 2018). Conversely, SDL focuses on effectively screening, managing, and utilizing massive data sets based on science research problems to resolve research issues and achieve innovative science research outcomes (Qin & D'ignazio, 2010).

Effective assessment tool for science data literacy
This study also developed SDL test questions based on the constructed conceptual framework, using physics, astronomy, geography, and other STEM subjects as examples, to assess the level of SDL among these specialized college students. The validity of the test questions was examined by the Rasch model, which is assumed to be unidimensional, which means that the test measures a single primary underlying feature (Weller, 2013). This study verified the unidimensionality of the test questions by analyzing the unexplained first eigenvalues and correlation ratios, confirming compliance with the Rasch model. Thus, the test instrument measured only one potential characteristic, SDL. Moreover, the study examined the reliability and separation of items and item fit at an overall level, and the results showed that the test questions had good reliability, separation, and item fit, proving the validity of the assessment tool.

The wright map was employed to assess the relative difficulty of test items in comparison to the ability values of the subjects, visually depicting the correspondence between items and participants (Glamočić et al., 2021). The results indicate variations in item difficulty, with some items being comparatively easier for students. Considering the grade distribution of participants, with over 70% being sophomores on the brink of entering their junior year and having undergone nearly two academic years of systematic science learning and inquiry practice, the test was generally perceived as slightly less challenging for them. Consequently, the SDL test questions developed in this study are deemed more suitable for assessing students who are new to college or at a lower grade level. To address higher-grade levels, a potential follow-up approach involves increasing the difficulty and contextual complexity of the test questions.

The DIF analysis based on gender reveals significant differences in some test items. Specifically, exercises falling under the categories of DAA, DIS, DCASS, and DLAR were found to be easier for girls, potentially attributed to their generally heightened perceptiveness and verbalization skills (Kan & Bulut, 2014). Conversely, exercises categorized as DVA and DSA were observed to be easier for boys. The prior survey reflected this trend, with more boys among the participants having prior experience with experiments related to science data and more exposure to data value and safety issues. The formation of gender differences may also be strongly influenced by cultural context. In numerous cultures, girls tend to be encouraged to develop language and social skills, while boys are more likely to be directed to participate in scientific and technical fields. This gender role stereotype may lead to girls performing more prominently in tasks involving perception and language, while boys are more dominant in tasks that involve processing data and conducting scientific experiments (Else-Quest, et al., 2010; Reilly et al., 2019a, 2019b). To enhance the cross-gender fairness of the test questions, we made careful adjustments to the question context and answer options to minimize the potential impact of cultural biases and gender stereotypes. These adjustments help ensure that the test fairly evaluates the abilities of all participants and is not limited to those individuals who conform to traditional gender role expectations. The revised question formulations are detailed in the Appendix.

It is noteworthy that the SDL assessment tool developed in this study shares certain similarities with existing DL assessment tools but also exhibits significant differences. First, a commonality lies in the format of the assessment tool, which, similar to some existing DL assessment tools such as the one devised by Pratama et al. (2020) for middle school students, takes the form of multiple-choice questions commonly used in DL assessments. Second, there is an overlap between the assessment tools in this study and existing DL tools concerning what is evaluated, with a shared focus on elements like accessing, managing, and analyzing data and communicating results (McGowan et al., 2022; Pratama et al., 2020). However, notable differences exist between this study's assessment tool and previous DL tools. First, the assessment tool in this study is built upon the conceptual framework of SDL, whereas existing DL assessment tools are founded on connotations related to DL, not specifically SDL (Pratama et al., 2020). Second, this study's assessment tool is comprehensive, concentrating on measuring various dimensions and elements of SDL, including science data awareness, science data skills, and science data regulations and ethics. In contrast,

existing DL measures primarily focus on data use, with limited attention to data regulations and ethical aspects and a lack of measurement of data awareness (McGowan et al., 2022; Trantham et al., 2021). Third, the assessment instrument in this study targets the SDL that college students in STEM-related majors should possess when faced with science data, as opposed to general data-handling skills or the use of data in specific professions (e.g., teachers) (Trantham et al., 2021).

Theoretical and practical value

The theoretical significance of this study is anchored in the development of a comprehensive conceptual framework for SDL, encompassing three dimensions: science data awareness, science data skills, and science data regulations and ethics. This framework not only clarifies the interrelationships and significance of these dimensions and their respective elements but also offers a more holistic understanding of SDL. It effectively addresses the shortcomings in prior definitions and interpretations of SDL, paving the way for a deeper appreciation and advancement of this field. Importantly, a nuanced comprehension of SDL and its determinants, particularly for STEM college students, equips educators and policymakers to more effectively tailor SDL development programs. This approach ensures the design and implementation of successful education strategies aimed at nurturing future talent with a high degree of SDL. In addition, enhancing students' SDL contributes to their overall science and DL, equipping them to navigate challenges related to big data and scientific issues they may encounter (Gebre, 2022; Sholikah & Pertiwi, 2021). Furthermore, our research underscores the need for heightened public awareness about the significance of SDL. In an era increasingly dominated by data-driven decision-making (Trantham et al., 2021), the general public's proficiency in SDL can significantly influence societal development and innovation. Our findings thus offer vital support for the community in recognizing and understanding the critical importance of SDL.

In practice, the SDL assessment tool created by this study is proficient in accurately measuring the level of SDL among STEM major college students in fields such as physics, astronomy, and geography. The results obtained from this tool can offer insightful guidance for crafting programs aimed at fostering SDL. Significantly, the study underscores the often neglected yet crucial role of SDL in previous research within this field. The difficulty faced by students in responding to two particular test items, item 5 and item 6, which focus on science data awareness, highlights the necessity of this quality, which is often found lacking among students. Therefore, when

enhancing SDL, it is essential to focus not only on developing science data skills but also on amplifying science data awareness. For instance, in educational environments such as classrooms, students should be encouraged to independently seek out relevant science data based on the requirements of the problem they are solving, rather than relying on pre-prepared data sets (Kjelvik et al., 2019; Schultheis & Kjelvik, 2020). This approach fosters both the awareness and proficiency of students in effectively utilizing science data.

It is worth noting that the SDL conceptual framework of this study was constructed based on an in-depth analysis of SL, DL, and science data characteristics. Therefore, the framework is applicable and universal not only to STEM subject areas but also to other scientific fields. The assessment test questions in this study are mainly applicable to undergraduates in the disciplines of physics, astronomy, and geography because some of the test questions have scientific contexts and science data related to these disciplines. However, these test questions can be adapted to students in different majors and at different levels of education with appropriate modifications. To accommodate different majors, test questions can be customized and adapted by selecting or designing science data and scenarios that match the characteristics of specific disciplines. For example, replacing test question contexts in physics and astronomy with chemistry lab-related contexts and using data from chemical disciplines such as chemical reactions would result in a more accurate assessment of chemistry majors' SDL. In addition, the difficulty of the test questions could be adjusted to target students at different educational levels. For example, for K-12 students, test questions can be made less difficult by simplifying the complexity of the science data, while for graduate students, test questions can be made more challenging by increasing the depth and breadth of the science data.

Conclusions and limitations

This study successfully developed a conceptual framework for SDL, featuring three dimensions and eleven elements, specifically designed for college students majoring in STEM. Furthermore, an SDL assessment tool was created based on this framework. This study addresses previous shortcomings, including the lack of a comprehensive definition and understanding of SDL, as well as the absence of suitable assessment tools. The outcomes of this research not only facilitate a deeper understanding and promotion of SDL but also provide essential support for nurturing SDL among college students, particularly those in STEM fields, and for enhancing public awareness of the importance of SDL.

However, there are certain limitations to this study. Initially, the assessment instrument for SDL employed real science data that were sourced from China and reflective of its context. While the choice of data was not contingent upon specific cultural knowledge but was instead aimed at evaluating students' general SDL skills, there is a possibility that students in other countries may be more engaged when the data relate to their own national contexts. We recommend and endorse the adaptation of the test questions by researchers in various countries, potentially by substituting the data with authentic examples that reflect their local or regional scientific landscapes. Second, we used some STEM majors, physics, astronomy, and geography, as examples for the development of the SDL assessment. Although the test questions are primarily focused on students' understanding and utilization of science data, it is important to acknowledge that future researchers could enrich and diversify the test questions by integrating their own disciplinary perspectives. Third, in response to the DIF findings, we refined the test questions, and we anticipate that future research will continue to investigate the fairness of these questions in terms of gender balance.

The limitations of this study reveal important directions for future research. First, future research could develop more cross-culturally adaptive assessment tools to broaden the scope of participation in the study and enhance assessment accuracy to ensure that students' SDL can be assessed fairly and effectively regardless of their national backgrounds. Second, future research could conduct international comparative studies to gain insight into the development of and differences in students' SDL across different educational systems and cultural contexts. Third, future research can refer to the assessment tools in this study to expand the subject areas involved in SDL assessment, including other STEM subjects such as biology and chemistry, as well as the humanities and social sciences. Fourth, future research can deepen the study of SDL assessment from the perspective of gender equality. In the design process of assessment tools, gender sensitivity is fully considered and questions that may trigger gender bias or stereotyping are avoided; pay attention to the causes of gender differences in SDL assessment results in practice and conduct more targeted training. Finally, future research could emphasize the development of students' SDL, and use assessment tools to measure the effectiveness of students' SDL development after receiving relevant educational interventions.

Qiao *et al. International Journal of STEM Education* (2024) 11:25

Page 18 of 21

Appendix: SDL assessment test questions

Dimension	Element	Item	Content introduction
Science data awareness	Data acuity awareness	1	Which of the following places can you find the first-hand data you need to know about earthquakes occurring across China?
		2	Ming wants to study recent astronomical phenomena and needs to collect a large number of picture resources, which of the following sources are most reliable?
	Data value awareness	3	Hong wants to build a model of the Temple of Heaven with blocks, which of the following data would be most helpful to Hong in purchasing the blocks?
		4	Hong wants to characterize the occurrence of solar eclipses in the last decade, which of the following data does she not need?
	Data security awareness	5	Hong is a graduate student who collects research data from her lab. Which of the following behaviors is best for protecting data security?
		6	Which of the following statements is true about the use of experimental data?
Science data skills	Data finding and collection skills	7	When conducting a database search, which is the optimal way to filter the results when too many are found?
		8	Ming wants to find atmospheric temperature data measured at the Wuhan station of the Meridian Project at the National Space Science Center, which of the following is most appropriate?

Dimension	Element	Item	Content introduction
	Data organization and management skills	9	In March 2022, Ming downloaded some data of atmospheric density, atmospheric temperature, and Sodium layer density measured by Beijing Yanqing Station. To carry out a follow-up analysis effectively, which of the following storage names should he use for this data?
		10	Multimodal data contains formats such as video, audio, pictures, and numbers, etc. A group has collected a large amount of multimodal data, and to easily share the multimodal data with subgroups on different tasks, which is the most appropriate way to save the data?
	Data processing and analysis skills	11	To measure students' satisfaction with a science course, Ming designed 10 questions to make a questionnaire, the following is the data he collected from the questionnaire, which of the following is inappropriate for data processing?
		12	To visualize millisecond pulsars, Ming processes and analyzes all of the pulsar data obtained from the pulsar database, which of the following operations is not appropriate?
	Data interpretation skills	13	The figure shows a graph of earthquake data in China over the years. The size of the red seismic dots in the graph represents the intensity of the earthquakes, which of the following statements is correct?

Dimension	Element	Item	Content introduction
		14	Hong wants to find out if there will be an anomaly in geomagnetism during a certain period (02/27/02 - 03/06/02) and she is given the geomagnetism for that period, which of the following statements is not correct?
	Data deduction skills	15	The average yield per unit area of three experimental varieties of cucumber A, B, and C for five consecutive years is shown in the table below, which is the better choice of cucumber to be grown by the farmer uncle?
		16	Hong wants to go from A to B. There will be a delay of 1 minute when she encounters a traffic light on the way. Based on the data that has been collected 4 options have been designed, which option should she choose now that she is in more of a hurry?
	Data communication and sharing skills	17	Ming has made a data image, as shown in the figure below. He wants to introduce the meaning of the image to others, which of the following is not required?
		18	The figure shows a visualization of China's carbon dioxide emissions from 1997 to 2012. To convince your classmates of China's success in saving energy and reducing emissions, what do you think is the most convincing of the following statement logic?

Dimension	Element	Item	Content introduction
Science data regulation and ethics	Data laws and regulations	19	Hong wants to use the data in a data center, but she does not have the download permission, which of the following is correct?
		20	Ming is a data manager in an ecology lab at a university. A data analysis company is particularly interested in the data on the ecological environment of Wuhan in recent years and wants to buy the relevant data. Faced with his request, which is the most appropriate thing for Ming to do?
	Data ethics	21	Which of the following violates the ethics of using science data?
		22	Which of the following statements about data ethics is not true?

Availability of data and materials
The data sets generated and analysed during the current study are not publicly available due the data is confidential, but are available from the corresponding author on reasonable request.

Qiao *et al. International Journal of STEM Education*   (2024) 11:25

Page 20 of 21

Declarations

Ethics approval and consent to participate

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

Australian Bureau of Statistics. (2010). *What is statistical literacy and why is it important to statistically literate?* (Feature Article). Tasmanian State and Regional Indicators (March). Retrieved September 20, 2022, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1307.6Feature+Article1+2009

Ball, A. (2012). *Review of data management lifecycle models.* Research Report, Bath: University of Bath, UK.

Bowler, L., Aronofsky, M., Milliken, G., & Acker, A. (2020). Teen engagements with data in an after-school data literacy programme at the public library. *Information Research An International Electronic Journal. 25*(4).

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123–134.

Carey, M., Grainger, P., & Christie, M. (2018). Preparing preservice teachers to be data literate: A Queensland case study. *Asia-Pacific Journal of Teacher Education, 46*(3), 267–278.

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal Libraries and the Academy, 11*(2), 629–657.

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review, 9*(2), 1–22.

Council of the European Union. (2018). Council Recommendation of 22 May 2018 on Key Competences for Lifelong Learning (No. OJ 2018/C 189/01). *Official Journal of the European Union.*

D'Ignazio, C., & Bhargava, R. (2015). Approaches to building big data literacy. *Bloomberg data for good exchange.*

Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., & De Laat, C. (2012). Addressing big data challenges for scientific data infrastructure. In *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings* (pp. 614–617). IEEE.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin, 136*(1), 103.

Fataliyev, T. K., & Mehdiyev, S. A. (2019). Integration of cyber-physical systems in e-science environment: state-of-the-art, problems and effective solutions. *International Journal of Modern Education and Computer Science, 11*(9), 35.

Faundeen, J., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., ... & Zolly, L. S. (2014). *The United States geological survey science data lifecycle model* (No. 2013-1265). US Geological Survey.

Fotopoulou, A. (2021). Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility. *Information, Communication & Society, 24*(11), 1640–1657.

Fox, C. M., & Bond, T. G. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science, 331*(6018), 705–708.

Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology, 53*(5), 1080–1095.

Gibson, P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany, 105*(12), 1953–1956.

Glamočić, D. S., Mešić, V., Neumann, K., Sušac, A., Boone, W. J., Aviani, I., & Grubelnik, V. (2021). Maintaining item banks with the Rasch model: An example from wave optics. *Physical Review Physics Education Research, 17*(1), 010105.

Government of Canada. (2021). *Choose Science: Government of Canada and STEM.* Retrieved from URL: https://ised-isde.canada.ca/site/choose-science/en/government-canada-and-stem#6

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record, 117*(4), 1–22.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9 Suppl), II28.

Kan, A., & Bulut, O. (2014). Examining the relationship between gender DIF and language complexity in mathematics assessments. *International Journal of Testing, 14*(3), 245–264.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering, 29*(10), 2318–2331.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)* (pp. 404). IEEE.

Kelling, S., Hochachka, W. M., Fink, D., Riedew, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: a new paradigm for biodiversity studies. *BioScience, 59*(7), 613–620.

Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation, 56*, 21–31.

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.

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science, 49*(1), 3–14.

Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008). First aid recommendations for psychosis: using the Delphi method to gain consensus between mental health consumers, carers, and clinicians. *Schizophrenia Bulletin, 34*(3), 435–443.

Law, H., & Morrison, A. P. (2014). Recovery in psychosis: a Delphi study with experts by experience. *Schizophrenia Bulletin, 40*, 1347–1355.

Liao, L. L., & Lai, I. J. (2017). Construction of nutrition literacy indicators for college students in Taiwan: a Delphi consensus study. *Journal of Nutrition Education and Behavior, 49*(9), 734–742.

Linacre, J. M. (2019). Winsteps® Rasch measurement computer program. Retrieved from https://www.winst.com

Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A., & Phatak, V. (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare, 2*(1), 4–10.

McGowan, B. S., Ekeigwe, A., & Clase, K. (2022). Designing and assessing a data literacy internship program for graduate health sciences students. *Journal of the Medical Library Association: JMLA, 110*(4), 501.

Meridian Project Data Center. (2023). Retrieved from https://data2.meridianproject.ac.cn/

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution, 27*(2), 85–93.

Mustafee, N., Bessis, N., Taylor, S. J., Hou, J., & Matthew, P. (2020). Co-citation analysis of literature in e-science and e-infrastructures. *Concurrency and Computation: Practice and Experience, 32*(9), e5620.

National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, contexts, and consequences.* Washington, DC: The National Academies Press. https://doi.org/10.17226/23595

O'Connor, J. P., Penney, D., Alfrey, L., Phillipson, S., Phillipson, S. N., & Jeanes, R. (2016). The development of the stereotypical attitudes HPE Scale. *Australian Journal of Teacher Education (online), 41*(7), 70–87.

OECD (2019), *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris. https://doi.org/10.1787/b25efab8-en.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). THE PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, 105906.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society, 21*(2), 419–437.

Pedaste, M., Kallas, K., & Baucal, A. (2023). Digital competence test for learning in schools: Development of items and scales. *Computers & Education, 203*, 104830.

Pratama, M. A., Lestari, D. P., Sari, W. K., Putri, T. S. Y., & Adiatmah, V. A. K. (2020). Data literacy assessment instrument for preparing 21 Cs literacy: preliminary study. *Journal of Physics Conference Series* (Vol. 1440, No. 1, p. 012085). IOP Publishing.

PRC General Office of the State Council. (2021). *Notice on the issuance of the Action Plan for Enhancing Scientific Literacy of the Whole Nation (2021–2035)* [Notice No. 2021-19]. Retrieved from URL: https://www.gov.cn/gongbao/content/2021/content_5623051.htm

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51–59.

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.

Quyên, ĐT. N. (2014). Developing university governance indicators and their weighting system using a modified Delphi method. *Procedia-Social and Behavioral Sciences, 141*, 828–833.

Reilly, D., Neumann, D. L., & Andrews, G. (2019a). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education, 49*(1), 25–50.

Reilly, D., Neumann, D. L., & Andrews, G. (2019b). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist, 74*(4), 445.

Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G. J., Felzer, B. S., McIntyre, N. E., & Weathers, K. C. (2014). Completing the data life cycle: Using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment, 12*(1), 24–30.

Sander, I. (2020). What is critical big data literacy and how can it be implemented? *Internet Policy Review, 9*(2), 1–22.

Schneegans, S. and S. Nair-Bedouelle (2021) Scientific literacy: an imperative for a complex world. In *UNESCO Science Report: The race against time for smarter development*, *2021*, 17. Schneegans, S.; Straza, T. and J. Lewis (eds). UNESCO Publishing: Paris.

Schultheis, E. H., & Kjelvik, M. K. (2020). Using messy, authentic data to promote data literacy & reveal the nature of science. *The American Biology Teacher, 82*(7), 439–446.

Shaffer, J. F., Ferguson, J., & Denaro, K. (2019). Use of the test of scientific literacy skills reveals that fundamental literacy is an important contributor to scientific literacy. *CBE Life Sciences Education, 18*(3), ar31.

Shields, M. (2004). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly, 28*(2/3), 6–11.

Sholikah, L., & Pertiwi, F. N. (2021). Analysis of science literacy ability of junior high school students based on Programme for International Student Assessement (PISA). *INSECTA: Integrative Science Education and Teaching Activity Journal, 2*(1), 95–104.

Siarova, H., Sternadel, D., & Szőnyi, E. (2019). *Research for CULT Committee—Science and Scientific Literacy as an Educational Challenge*. European Parliament, Policy Department for Structural and Cohesion Policies, Brussels.

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research, 6*(1), 1–21.

Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). T. Hey (Ed.). Redmond, WA: Microsoft research.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE, 6*(6), e01.

Trantham, P. S., Sikorski, J., de Ayala, R. J., & Doll, B. (2021). An item response theory and Rasch analysis of the NUDKS: a data literacy scale. *Educational Assessment, Evaluation and Accountability*, 1–23.

Vahey, P., Yarnall, L., Patton, C., Zalles, D., & Swan, K. (2006, April). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In *Annual Meeting of the American Educational Research Association, San Francisco, CA* (pp. 1–15).

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*(2), 198–212.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), 1–9.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3).

Yang, S., Fichman, P., Zhu, X., Sanfilippo, M., Li, S., & Fleischmann, K. R. (2020). The use of ICT during COVID-19. *Proceedings of the Association for Information Science and Technology, 57*(1), e297.

Young, J. W., Morgan, R., Rybinski, P., Steinberg, J., & Wang, Y. (2013). Assessing the test information function and differential item functioning for the TOEFL Junior® Standard Test. *ETS Research Report Series, 3*(1), i–27.

Publisher's Note