Educational Research Review 52 (2026) 100828

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Review

A systematic review of frameworks and measurement instruments for data literacy in higher and adult education

Marlene Ganz [*], Marco Kalz

Heidelberg University of Education, Institute for Arts, Music and Media, Faculty of Cultural Science and Humanities, Im Neuenheimer Feld 561, 69120, Germany

A B S T R A C T

The concept of data literacy suffers from a persistent theoretical and definitional fragmentation. To contribute to more coherence in the field, we introduce two systematic literature reviews. Together, they provide an integrated overview of how data literacy is conceptualized in competency frameworks and represented in measurement instruments. The first review analyzed 15 frameworks developed for higher education and related contexts after applying rigorous inclusion and exclusion criteria. The content analysis focused on target groups, structural characteristics, and competency dimensions. The second review examined six instruments designed to assess data literacy after applying rigorous inclusion and exclusion criteria. The content analysis focused on targeted competencies, validation approaches, and provided a quality appraisal. The synthesis shows that frameworks describe data literacy in broad and multidimensional terms, whereas instruments capture only a narrower subset of competencies, often emphasizing technical skills while reflective and ethical dimensions remain underrepresented. Most instruments rely on self-report formats and provide limited validation evidence, raising concerns about their ability to capture applied and context-dependent competencies. Building on these insights, the study proposes a Core–Periphery Framework (COPERDALE) that distinguishes a stable core of fundamental skills from more flexible, context-dependent competencies. This structure clarifies which dimensions can be universally assessed and which require contextual adaptation. Beyond conceptual clarification, the findings underscore the need to diversify assessment practices, moving beyond self-reports toward approaches that better capture applied, reflective and ethical aspects of data literacy. The review thereby offers both a structured conceptual foundation and directions for advancing research and educational practice.

1. Introduction

Data literacy is a key qualification in the data-driven world of science, business and society. It encompasses skills such as analyzing, interpreting and applying data as well as understanding its ethical implications. Universities play a crucial role in preparing students to meet these demands. To develop targeted educational programs and evaluate their effectiveness, it is essential to measure students' data literacy as a quantifiable variable.

Although interest in data literacy has increased substantially in recent years, there remains considerable ambiguity regarding which frameworks, assessment tools and methodological approaches are most appropriate for evaluating data literacy in higher education.

Existing reviews highlight inconsistencies in defining core competencies. Empirical research identifying competencies relevant to specific target groups remains limited, while a broadly accepted understanding of data literacy that spans across disciplinary boundaries is still lacking (Ghodoosi et al., 2023). Furthermore, the current literature offers little conceptual clarity on how data literacy should be understood and operationalized within educational settings, particularly at the tertiary level (Lee et al., 2024). In addition, there is limited evidence on the validation and contextual adaptation of assessment tools. Many of the existing tools have

* Corresponding author. Heidelberg University of Education, Im Neuenheimer Feld 561, 69120, Germany.
E-mail address: ganz@ph-heidelberg.de (M. Ganz).

https://doi.org/10.1016/j.edurev.2026.100828
Received 18 August 2025; Received in revised form 25 July 2026; Accepted 27 July 2026
Available online 28 July 2026
1747-938X/

been developed within narrowly defined domains or for specific interventions and provide only preliminary or fragmentary information on their psychometric quality (Cui et al., 2023).

In light of these gaps, there is a pressing need for a systematic investigation that not only maps existing definitions and competencies but also evaluates how data literacy is being assessed in practice, especially within higher education.

This review addresses this need by providing an integrated analysis of two closely connected areas: (1) frameworks that define data literacy competencies and (2) measurement tools and methodologies used to assess them. This dual approach highlights the diversity of conceptualizations and assessment strategies. It also identifies methodological strengths and limitations and reveals gaps in the current research landscape. While the primary focus lies on higher education, the analysis also considers relevant models and instruments from adjacent domains where transferable insights can be drawn. This broader perspective provides a more comprehensive picture of the field.

The findings aim to support the development of a tailored competence model and an adapted assessment tool for use in both future research and educational practice.

To guide this synthesis, the review explores the following overarching question:

1. Which conceptual dimensions of data literacy are reflected in current frameworks and assessment tools?
2. To what extent is their empirical validation supported by existing evidence?

1.1. Conceptual foundation

1.1.1. Defining data literacy

Even though data literacy is seen as a fundamental skill for dealing with complex social and professional challenges, a universally valid definition remains difficult to find, as interpretations vary depending on the discipline and context.

Wolff et al. (2016) conducted an extensive qualitative analysis of existing definitions to identify commonalities and propose a unified perspective. Their findings revealed recurring themes, including the importance of real-world contexts, critical-thinking skills and the iterative process of data inquiry. This process encompasses stages such as problem formulation, data collection, analysis, interpretation and decision-making. Wolff et al. also stressed that data literacy extends beyond technical skills to include ethical considerations and the ability to adapt to various user scenarios. Based on these insights, they proposed the following definition:

> “Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyze, visualize, critique, and interpret data, as well as to communicate stories from data and to use data as part of a design process" (Wolff et al. 2016, p. 23).

The term data literacy is also often used interchangeably with related concepts such as information literacy, media literacy and digital literacy, which can lead to confusion. While these concepts share similarities, they address distinct aspects of working with information and technology. According to the American Library Association (1989), “to be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information” (p. 1). Media literacy, on the other hand, is defined as “the ability to access, analyze, evaluate and create messages across a variety of contexts” (Livingstone, 2004, p. 3). This four-part model applies to both traditional and digital media and emphasizes active participation in media environments. Digital literacy, as conceptualized by Eshet (2004), goes beyond basic technical skills and includes a variety of cognitive, sociological and emotional abilities required to operate successfully in digital environments.

While information, media and digital literacy address broader competencies in working with information and technology, data literacy focuses explicitly on competencies required to analyze, interpret and ethically use data in diverse contexts (Deahl, 2014; Mandinach & Gummer, 2013; Wolff et al., 2016). At the same time, data literacy may draw upon these adjacent literacies or even rely on them as prerequisites. Skills such as navigating digital tools, assessing the credibility of information or understanding media contexts can form a foundational basis for meaningful data engagement. This is reflected in earlier definitions of data literacy, which were closely linked to concepts like information literacy and have since been extended to include data-specific competencies such as analysis and interpretation (Bhargava & D'Ignazio, 2015). In contrast, statistical literacy emphasizes understanding statistical concepts and critically evaluating quantitative information (Gal, 2002; Schield, 2004). Although data literacy shares this quantitative focus, it tends to encompass a broader spectrum of skills, including ethical reasoning, contextual awareness and the communication of insights derived from data (Deahl, 2014; Mandinach & Gummer, 2013; Wolff et al., 2016). From this perspective, statistical literacy can be seen as a valuable subset of data literacy, yet not sufficient to capture its full complexity.

While these comparisons help to clarify the conceptual landscape surrounding data literacy, they also underscore its unique position as a cross-cutting and evolving set of skills. As Wolff et al. (2016) highlight, the growing use of data in everyday decision-making requires a dynamic and adaptable set of competencies. Given its diverse dimensions and the rapid development of technologies, data literacy could be understood as a multifaceted construct encompassing technical, ethical and interpretive competencies. It may also be possible that data literacy is not an independent competence, but rather a combination of various other competencies, depending on the context and specific tasks involved.

1.1.2. Conceptualizing competency frameworks

In the literature, the terms “competency model” and “framework” are often used interchangeably. However, the term competency model can sometimes be misleading, as many of these models primarily serve as compilations or listings of competencies rather than as structured models. According to Krumm et al. (2012), in such cases, the terms competency framework or competency profile might be more accurate. Nevertheless, he emphasizes that competency model remains the more commonly used term in practice.

Competency models are not only fundamental for developing educational curricula (Niegemann et al., 2008) but also for assessing and measuring competencies (Klieme & Leutner, 2006; Shavelson, 2010; Shavelson et al., 2005). They provide structured models that define and organize key elements of a domain, competency or learning objective (Shavelson et al., 2005). According to Niegemann et al. (2008), such models delineate complex competencies by specifying their sub-competencies and interrelationships. Similarly, Krumm et al. (2012) emphasize that competency models function as structured compilations and descriptions of relevant competencies necessary for effective performance in a given context.

Klieme and Leutner (2006) highlight that structured competency models serve as essential tools for assessing individual learning outcomes by systematically categorizing different types of knowledge and skills. They also contribute to making teaching objectives explicit by organizing knowledge into distinct dimensions, such as declarative, procedural, schematic and strategic knowledge (Shavelson et al., 2005). These classifications facilitate the translation of learning objectives into measurable competencies, forming a foundation for assessment. Glaser et al. (2001) stress the importance of establishing clear connections between what is taught, how it is taught and how learning is assessed. By linking objectives to observable and assessable performance indicators, frameworks assist educators in developing valid and reliable assessment tools that reflect students' competencies and progress.

Due to the lack of standardized guidelines for categorizing knowledge or formulating learning objectives, competency model design varies depending on the application context. In this regard, Klieme (2004) highlights the importance of empirically grounded models that should be developed inter disciplinarily and tailored to specific contexts.

Some models structure competencies by dividing them into different sub-competencies and their interrelations (Fleischer et al., 2010). Other models define proficiency levels to describe varying degrees of (sub-)competency development (Klieme, 2004; Klieme & Leutner, 2006).

While many educational frameworks, such as those described by Shavelson et al. (2005), emphasize structured categorization and assessment, Klieme and Leutner (2006) argue that competency models must be flexible enough to serve both individual diagnostic purposes and large-scale educational assessments. Unlike rigid models with fixed competency levels, some frameworks enable context-dependent adaptation, ensuring the applicability of knowledge across diverse learning environments.

For the purpose of this review, a competency framework is understood as a structured representation of competencies that systematically defines, categorizes and interrelates skills and knowledge within a specific domain. Such frameworks typically include dimensions, sub-competencies or proficiency levels that illustrate different degrees of competency development. They serve as conceptual tools for structuring competencies and may be used in educational, professional or assessment contexts.

Competency frameworks differ from simple lists or compilations of competencies in that they provide an explicit structure that organizes competencies into meaningful categories and relationships. While some frameworks establish fixed proficiency levels or follow taxonomic structures, others are more flexible, allowing for adaptation to specific contexts or evolving requirements.

2. Shared methodology

This study consists of two systematic literature reviews with distinct but related objectives: one focuses on frameworks for data literacy competencies, the other on measurement instruments. Despite their different focal points, both reviews share a common methodological foundation that ensures alignment and consistency throughout the entire study.

2.1. Review design and reporting standards

Both reviews were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model and based on established recommendations for high-quality systematic reviews (Moher et al., 2010; Page et al., 2021; Polanin et al., 2019; Wolfswinkel et al., 2013). The PRISMA 2020 guidelines served as a structured framework to ensure transparency, replicability and methodological rigor across all stages of the review process. The PRISMA checklist and flow diagram were used to document the identification, screening, eligibility assessment and inclusion of studies (Moher et al., 2010; Page et al., 2021).

2.2. Category system for competency mapping

To enable a systematic analysis and comparison of the competency dimensions addressed in data literacy frameworks and measurement instruments, a category system was developed as part of the first review on data literacy frameworks and subsequently also applied to the second review on assessment instruments. The development of this system followed a multi-step, iterative process conducted during the analysis of the frameworks.

The starting point for the analysis was the set of the most highly cited articles available at the time the review was initiated. Wolff et al. (2016) synthesized key data literacy competences into a coherent model structured around the PPDAC inquiry cycle (Problem – Plan – Data – Analysis – Conclusion) (Wild & Pfannkuch, 1999). In addition to core skills like asking questions, analyzing and interpreting data, their framework also addressed overarching dimensions such as ethics and real-world relevance.

Building on additional highly cited literature (Deahl, 2014; Carlson et al., 2011; Mandinach & Gummer, 2013), further competences such as communication & storytelling and data management were added, which were not explicitly addressed in Wolff et al.'s model. The identified competences were initially grouped into 14 categories, each accompanied by a concise definition to clarify its scope. These categories formed the deductive core of the system.

During the mapping of frameworks, additional categories were developed inductively whenever a competence appeared in at least two frameworks but was not covered by the existing categories. This resulted in a final set of 21 dimensions, including 14 deductively developed and seven inductively added categories.

The final category system was then applied unchanged to the analysis of measurement instruments. In both reviews a competence was coded as present if it was explicitly described or clearly implied in the respective source.

3. Systematic review of data literacy frameworks

3.1. Research questions

To explore how data literacy is conceptualized, this review focuses on existing frameworks that define relevant competencies and inform educational practice. The analysis is structured around the following research questions:

RQ1: What data literacy frameworks exist, and for which target groups have they been developed?
RQ2: What key components or competency dimensions of data literacy are included in these frameworks?
RQ3: How do frameworks differ or overlap in their structure and focus?

3.2. Data sources and search strategy

To identify relevant articles on data literacy frameworks, a comprehensive search was conducted across eight academic databases between 2 and 7 February 2025. The databases were selected for their relevance to educational research, interdisciplinary studies, and technology-related publications. The following databases were searched: ERIC, EBSCOhost, Web of Science, Springer Link, Taylor & Francis Online, IEEE Xplore, and BASE. In addition, Google Scholar was used to broaden the search and identify further relevant academic and grey literature.

We have followed PRIMSA recommendations (Page et al., 2021) and a single author has conducted the search process. We have mitigated the potential problem of missing important studies in this stage by applying multiple search strategies (snowball search, AI based search).

The search strategy was developed iteratively through exploratory searches across multiple databases. Initial attempts using combinations of the terms “data literacy” and “framework”, including their German equivalents “Datenkompetenz” and “Kompetenzrahmen”, revealed that title and keyword searches often yielded few or irrelevant results, particularly for the term “framework,” which frequently appeared only in full texts rather than in titles or metadata. Broader alternatives such as “data skills,” “data competence,” or “model” were tested but proved too imprecise. While these terms increased the number of hits substantially, sample-based relevance checks showed that they introduced considerable noise with little added value.

The most effective approach combined “data literacy” or “Datenkompetenz” in the title with “framework”, “Kompetenzrahmen”, “competence model” or “Kompetenzmodell” in all searchable fields. Omitting contextual terms such as “higher education” broadened the scope while still retrieving studies, particularly from teacher education, that were clearly situated in academic contexts. These frameworks may provide transferable insights for higher education.

All search strings were adapted to the syntax and field structures of each database. The complete list of search strings used across all databases is provided in the appendix (Table A. 1).

Search terms in all databases followed a consistent conceptual structure but were adjusted to each platform's technical constraints (e.g., Springer Link allows only limited filtering options). No filters were applied for time period or publication type, although only

Table 1
Inclusion and exclusion criteria by category – frameworks.

Category	Inclusion Criteria	Exclusion Criteria
Framework	Developed a new framework or extended an existing one for data literacy	Offered purely theoretical discussions or unspecified models without a clear structure or defined components
Conceptual Orientation	Reflected a broad, practice-oriented understanding of data literacy, consistent with perspectives that emphasize the ability to analyze, interpret and ethically use data, often through iterative inquiry and critical reflection in diverse contexts	Did not reflect a broad, practice-oriented understanding of data literacy and lacked consideration of analytical, interpretive or ethical use of data, including iterative inquiry and critical reflection across diverse contexts
Structural Clarity	Proposed, refined or conceptualized a structured representation of competencies that systematically defines, categorizes and interrelates relevant skills and knowledge within a specific domain	Presented unstructured compilations or lists of competencies lacking systematic definition, categorization, or internal relationships
Publication Type	Peer-reviewed journal articles, conference papers, book chapters, selected grey literature (e.g., institutional reports)	Commentary or opinion pieces
Time Scope	Any	-
Language/ Geographical Scope	Written in English or German, any geographical origin	Written in languages other than English or German

articles in English or German were included. Google Scholar was used as an additional source through focused title searches, given its broad indexing and limited filtering capabilities.

This approach ensured a conceptually focused yet inclusive search strategy suitable for identifying relevant frameworks across educational domains. To ensure ongoing coverage, search alerts were set up in Google Scholar and other platforms to capture newly available publications throughout the review process.

3.3. Selection process

The selection process followed a multi-stage procedure aligned with the PRISMA 2020 guidelines. In the first step, all records identified through the database searches were imported into a reference management tool. After the removal of duplicates (n = 154), a total of n = 271 records remained for screening.

The title and abstract of each record were screened based on the previously defined inclusion and exclusion criteria. At this stage, articles were excluded if they clearly did not meet the basic topical criteria (e.g. not related to data literacy, not focused on frameworks). This step reduced the pool to n = 58 records deemed potentially relevant for full-text review.

In the second screening phase, the full texts of these n = 58 articles were examined in detail. The applied inclusion and exclusion criteria are summarized in Table 1. This step focused on identifying studies that presented a concrete data literacy framework or model and appeared to meet the structural and thematic requirements defined for inclusion.

Following this phase, a final critical appraisal was conducted for studies where uncertainties remained regarding their compliance with the framework-specific criteria established for this review (n = 9). Frameworks that, upon closer examination, did not sufficiently fulfill the structural or conceptual requirements were excluded. In some cases, the distinction was not straightforward, as studies showed certain characteristics of competency frameworks or explicitly referred to their approach as a “framework”, while not fully meeting the definition adopted for this review. Borderline cases included, for example, studies that identified and thematically grouped relevant data literacy competencies but did not organize them into a systematic competency framework, as well as studies that proposed a framework primarily as a didactical approach to fostering data literacy rather than as a structured representation of data literacy competencies. During the full-text assessment, it also became apparent that data literacy was conceptualized with different orientations across the literature. Frameworks with a primary focus outside the conceptual scope of the review, such as approaches centered predominantly on personal data, data rights, and civic engagement, were therefore not included in the synthesis. In total, n = 10 studies were retained for inclusion, following a transparent selection process that included detailed discussion of all studies with

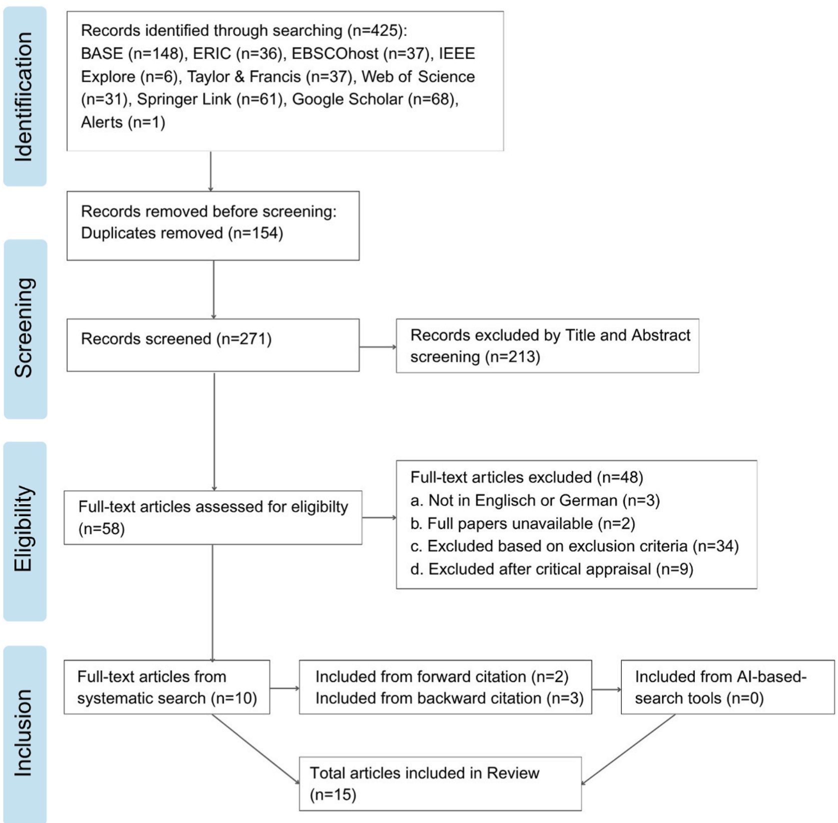

Fig. 1. PRISMA flow diagram of the literature selection process – frameworks.

remaining uncertainties.

In addition, a systematic backward and forward snowballing procedure was conducted by manually screening the reference lists and citations of all studies that passed the full-text screening. Screening was primarily based on the titles of referenced sources, with contextual cues in the main text also considered when references appeared related to data literacy frameworks. This process yielded five additional records that met the predefined criteria. An additional AI-based search was also conducted but did not identify any further relevant studies.

An initial set of three randomly chosen articles from the included framework sample was used to test the interpretation overlap of both authors. This selection led to a perfect overlap in interpretation pointing to a high interrater reliability. Further screening steps were conducted by a single reviewer. Conflicting cases were discussed in multiple consultation sessions of both authors. Final selection of included and excluded articles was discussed in consultation with a second researcher to ensure transparency and consistency in the application of the predefined criteria. To ensure transparency, reasons for inclusion or exclusion were documented at each stage of the screening process. This procedure is visualized in the PRISMA flow diagram (see Fig. 1).

3.4. Results

3.4.1. Frameworks and target groups

The reviewed frameworks address a diverse range of target groups. Most frameworks were developed for use in higher education (n = 8), targeting university students across disciplines (Bush et al., 2021; Calzada Prado & Marzal, 2013; Echtenbruck et al., 2025; Ridsdale et al., 2015; Schüller et al., 2019) as well as students in specific subject areas such as STEM, translation studies or public administration (Qiao et al. 2024; Hackenbuchner & Krüger, 2023; Overton & Kleinschmit, 2022). These frameworks often aim to enhance students' employability, academic success and civic engagement or to inform curriculum development within the respective disciplines. In addition, some frameworks include academic staff, researchers or institutional stakeholders as part of their intended audience (Calzada Prado & Marzal, 2013; Echtenbruck et al., 2025; Ridsdale et al., 2015; Schüller et al., 2019). Their purpose is typically broader, aiming to support institutional capacity-building and the strategic integration of data literacy into teaching and research practices.

Several frameworks focus on teacher education (Mandinach & Gummer, 2016; Raffaghelli, 2019), addressing the role of data use in pedagogical practice and teacher professional development across all education levels. Closely related are frameworks for educators in digital and blended learning environments, which emphasize instructional design and the use of learning analytics for personalized feedback and course improvement (Papamitsiou et al., 2021).

One frameworks target K–12 students, proposes a model for teaching data literacy in school curricula and fostering critical engagement with data from an early age (Grillenberger & Romeike). However, the authors emphasize the transferability of the framework to higher education contexts.

Table 2
Application contexts and target groups of included frameworks.

	Application Context	Author (Year)	Target Group
1	Higher Education Context	Bush et al. (2021)	University students across all disciplines
2		Calzada Prado & Marzal (2013)	University students, also relevant for academic library staff and instructors
3		Ridsdale et al. (2015)	University students, also relevant for researchers, professionals, and policymakers
4		Schüller et al. (2019)	University students, also relevant for educators and institutions
5		Echtenbruck et al. (2025)	University students across all disciplines, also relevant for lecturers and researchers in higher education
6	University Students (Discipline-Specific)	Qiao et al. (2024)	University students in STEM disciplines
7		Hackenbuchner & Krüger (2023)	University students in translation studies programs
8		Overton & Kleinschmit (2022)	University students (Public Administration)
9	Teacher Education	Mandinach & Gummer (2016)	School teachers in K–12 education
10		Raffaghelli (2019)	Educators from early education to higher education and adult learning
11	Educators (digital & blended learning contexts)	Papamitsiou et al. (2021)	Instructional designers and e-tutors in online and blended learning contexts
12	Students (K-12)	Grillenberger & Romeike (2018)	Secondary school students; transferable to other educational levels
13	Education Policy	Kim & Kang (2024)	Policy-makers, curriculum developers and the general public
14	Public Sector Context	Ongena (2023)	Public sector professionals and administrative staff
15		Government of Canada (2023)	Public servants at all levels and functions within the Government of Canada

Note. Application Context refers to the primary educational or professional setting for which the framework was developed. *Target Group* refers to the primary intended users of the framework as described by the original authors. Where applicable, additional intended user groups explicitly identified in the original publication are also reported.

Beyond the educational field, some frameworks are situated in the public sector, targeting government employees and public servants at different administrative levels (Ongena, 2023; Government of Canada, 2023). They aim to strengthen data-informed decision-making, promote responsible and critical data practices and support strategic workforce development in public administration.

Lastly, one framework (Kim & Kang, 2024) explicitly addresses policy-makers and the general public, with the purpose of guiding national education strategies and raising broad awareness of data literacy as a key competence in the digital age.

Overall, most frameworks were developed for use in educational settings, particularly in higher education. In addition, several frameworks address contexts beyond academia, including public sector institutions and policy development. An overview of the application contexts and corresponding target groups is provided in Table 2. Further descriptive information on the included frameworks is provided in Table B. 1 (Appendix B).

3.4.2. Key dimensions and structural characteristics of frameworks

3.4.2.1. Shared and divergent competency dimensions across frameworks. To address the research questions regarding the key components of data literacy (RQ2) and structural similarities and differences across frameworks (RQ3), the presence of 21 competency dimensions was examined across all included frameworks. These dimensions were derived through a combined deductive–inductive coding process. The mapping results are presented in a heatmap (Fig. 2), which displays the presence or absence of each dimension in each framework.

The competency dimensions most frequently included across frameworks are Data Analysis and Data Interpretation, which are represented in all 15 frameworks (100%), followed by Identifying & Accessing Data, Data Collection, Data Preparation, Data Visualizing, and Ethics, each appearing in 14 of the 15 frameworks (93%).

Beyond technical core skills, several frameworks include dimensions that address the critical and contextual understanding of data. These competencies are less universally represented and their distribution shows that the more analytically demanding or reflective a competence is, the less frequently it appears across frameworks.

A relatively common category is *Understanding Data*, included in 10 of the 15 frameworks (67%). It reflects a conceptual awareness of what data is, how it is generated and how it operates in different contexts. *Evaluating Data*, present in nine frameworks (60%), builds on this by encouraging a more critical appraisal of data quality, origin and processing. Only four frameworks (27%) address *Evaluating Data Analyses*, which involves reconstructing and critically assessing analytical procedures in light of the underlying data. Finally, *Assessing Societal* Implications, found in just two frameworks (13%), refers to the ability to reflect on the broader societal and organizational consequences of data use, both positive and negative, for individuals, communities and institutions.

Communication & Storytelling and *Data-Driven Action* appear in 10 of the 15 frameworks (67%). Both dimensions focus on how insights gained from data are put to use. *Communication & Storytelling* refers to the ability to support arguments and convey insights clearly and effectively through data. In contrast, *Data-Driven Action* emphasizes turning those insights into concrete strategies and decisions.

Data Sharing & Citation and *Data Culture* are addressed in eight of the 15 frameworks (53%). Both dimensions illustrate the role of responsible and collaborative data practices. *Data Sharing & Citation* involves the legal, ethical and technically appropriate sharing of data, including proper documentation and attribution. *Data Culture* refers to the ability to recognize the value of data and to contribute

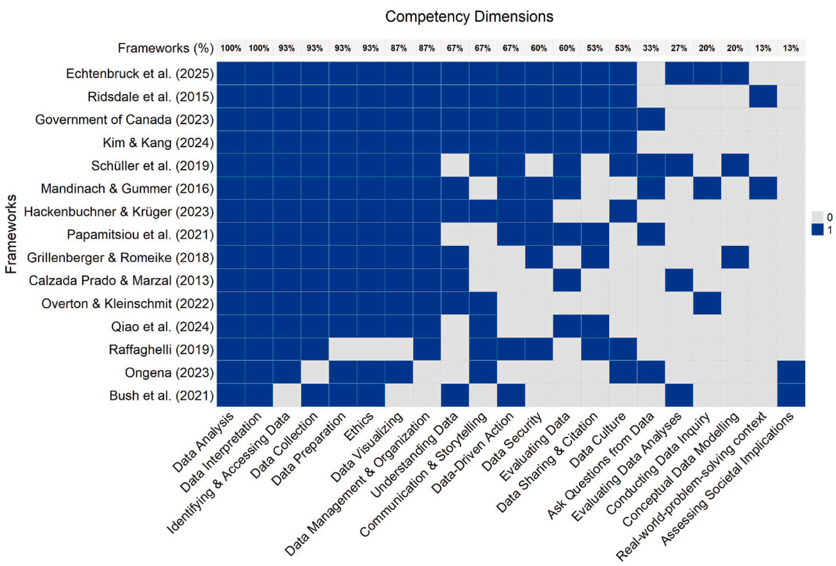

Fig. 2. Heatmap of competency dimensions covered across reviewed frameworks.

to an environment that supports responsible, reflective and strategic use of data.

The categories *Ask Questions from Data, Conducting Data Inquiry, Real-World-Problem-Solving Context,* and *Conceptual Data Modeling* reflect the framing and application of data literacy within authentic, inquiry-driven processes. The ability to formulate relevant and researchable questions (*Ask Questions from Data*) appears in five of the 15 frameworks (33%) and serves as a foundational entry point for data work. *Conducting Data Inquiry*, which encompasses the structured coordination of the entire data process, is only found in three frameworks (20%). *Real-World-Problem-Solving Context*, which refers to using data to explore real-world issues, is present in two frameworks (13%). *Conceptual Data Modeling*, which involves abstracting real-world phenomena into structured data representations, also appears in only three frameworks (20%) and represents a more specialized competence within the inquiry process.

The categories *Data Management & Organization* and *Data Security* refer to structuring and protecting data throughout its lifecycle. *Data Management & Organization* is one of the most frequently included dimensions, appearing in 13 of the 15 frameworks (87%). *Data Security*, addressed in nine frameworks (60%), refers to competencies required to safeguard data from unauthorized access, ensure confidentiality and comply with relevant legal and institutional standards.

In addition to variation across dimensions, the heatmap also reveals considerable differences in overall scope between the frameworks. Some frameworks address almost all competency dimensions, while others are highly selective in their focus.

3.4.2.2. Contrasts across frameworks. Among the included frameworks, four stand out for their breadth and balance in covering a wide range of competency dimensions: Ridsdale et al. (2015), Kim & Kang (2024), Government of Canada (2023) and Echtenbruck et al. (2025). Despite originating from different contexts, higher education, public administration and education policy, three of them (Ridsdale et al., 2015; Kim & Kang, 2024; Government of Canada, 2023) address 16 of the 21 categories, including nearly all of the most frequently observed dimensions such as *Data Analysis, Data Interpretation, Data Visualization, Ethics, Data Management & Organization,* and *Communication & Storytelling*. Their profiles are almost identical, differing only in one category. The framework by Echtenbruck et al. (2025) goes even further, addressing 18 of the 21 categories and thus offering the most comprehensive coverage among all reviewed frameworks. In addition to covering most technical and procedural dimensions, it includes research-oriented competencies such as Research Design, Open Science practices, and Copyright, which addresses broader issues of intellectual property rights and licensing in research publishing. Overall, reflective or critically evaluative elements are less prominently represented in these otherwise comprehensive models.

In contrast to the broadly structured frameworks, the model by Bush et al. (2021), designed for university students across disciplines, covers only nine of the 21 dimensions. However, it stands out for its strong emphasis on critical and reflective aspects of data literacy. The framework includes competencies such as *Evaluating Data Analyses*, *Ethics,* and *Assessing Societal Implications*, while omitting many of the technical and procedural categories. Beyond these codable categories, the framework places a particularly strong focus on critical reflection, ethical reasoning, and socio-political awareness. Several learning objectives explicitly stress the importance of considering the social, cultural and disciplinary context of data. The framework also draws attention to the ability to switch between disciplinary and societal perspectives and to engage with the implications of data use across different domains.

The framework by Raffaghelli (2019), situated in the field of teacher education, also covers only 10 of the 21 competency dimensions and is characterized by a specific pedagogical focus. Rather than addressing data-related skills in a general or procedural manner, it emphasizes data-informed teaching practices, collaborative professional learning, and institutional engagement with data. Notably, the framework highlights the role of data in promoting inclusive, learner-centered and reflective educational practices. It also stresses the importance of empowering both educators and students to critically engage with data in their respective roles.

A markedly technical and data science–oriented approach to data literacy offers the framework by Overton & Kleinschmit (2022), developed for students in public administration programs. It covers 15 of the 21 competency dimensions, with a strong emphasis on data processing, modeling, programming and reproducibility. In contrast to the more reflective or pedagogical frameworks, Overton & Kleinschmit prioritize procedural and computational aspects such as literate programming, version control and workflow organization. Ethical and legal issues are acknowledged but remain secondary.

Taken together, these examples illustrate the wide range of conceptual approaches to data literacy. Yet despite this diversity, no consistent pattern emerges across application contexts, as even frameworks within the same domain vary significantly in scope and focus.

3.4.3. Discussion of framework findings

The review of data literacy frameworks reveals both convergence and diversity in how data-related competencies are conceptualized across educational and institutional contexts. A shared foundation of technical and procedural competencies emerges, including *Identifying & Accessing Data, Data Collection, Data Preparation, Data Analysis, Data Interpretation, Data Visualizing, Data Management & Organization*. These competencies are present in the majority of frameworks and reflect a widespread understanding of data literacy as a practical, skill-oriented ability to work with data across its lifecycle. In addition, *Ethics* appears in nearly all frameworks, pointing to a broad recognition of responsible data use as a key component of data literacy.

Several frameworks also include dimensions such as *Communication & Storytelling* and *Data-Driven Action*, suggesting that data literacy often is understood to encompass not only analysis but also the communication and application of insights.

While a common set of core competencies is present in most frameworks, the analysis also reveals variation in scope and emphasis. Only a small number of frameworks exhibit a clearly identifiable focus, either by strongly foregrounding critical reflection and societal implications (Bush et al., 2021) or by embedding data literacy into pedagogical practice and institutional development (Raffaghelli, 2019). The majority of frameworks, however, remain without a strong thematic or structural orientation. A more specific focus is primarily found in discipline-oriented frameworks designed for particular domains such as STEM education, public administration or computer science (Qiao et al. 2024; Overton & Kleinschmit, 2022; Grillenberger & Romeike, 2018). These models often reflect the data practices and requirements of their respective fields, suggesting a contextual adaptation of data literacy competencies. For example, the framework by Overton & Kleinschmit (2022) places strong emphasis on procedural and computational skills and can be interpreted as conceptualizing data literacy primarily as a set of advanced, domain-general skills for research and applied contexts.

The intended target groups of the frameworks influence how data literacy is framed and what purposes it is meant to serve. Frameworks developed for higher education students or the general public tend to emphasize transferable skills (e.g., Kim & Kang, 2024; Ridsdale et al., 2015; Schüller et al., 2019) or, in the case of Echtenbruck et al. (2025), stand out for their comprehensive coverage and strong research orientation, whereas frameworks aimed at educators, teacher training or public administration embed data literacy more explicitly in professional roles and responsibilities (Mandinach & Gummer, 2016; Raffaghelli, 2019; Ongena, 2023). This highlights the context-dependent nature of data literacy and raises the question of whether a universal definition is feasible or even desirable.

Only one framework (Bush et al., 2021) places sustained emphasis on social critique, power dynamics and the ability to shift between disciplinary and societal perspectives. These aspects reinforce the framework's orientation toward a transformative and critically engaged understanding of data literacy. In this sense, Bush et al. (2021) conceptualize data literacy not merely as a set of skills, but as a form of civic and epistemic responsibility. While evaluative or reflective competencies appear in other models, they are typically embedded within broader procedural structures rather than forming a central orientation. This suggests that approaches to data literacy that center civic responsibility or critical engagement remain largely underrepresented in the current landscape.

Taken together, the examined frameworks illustrate a broad conceptual diversity in how data literacy is defined and operationalized. The findings point to a relatively stable core of data-related competencies that are widely recognized across different frameworks. At the same time, the overall field remains fragmented, with substantial variation in emphasis, scope and contextual grounding. This structural variation illustrates the heterogeneous ways in which data literacy is operationalized. While some frameworks reflect domain-specific needs or pedagogical aims, most refrain from articulating a distinct conceptual stance beyond procedural skill development. These observations raise important questions about whether current frameworks fully reflect the conceptual and practical breadth of data literacy as it is needed in education, research and society.

4. Systematic review of data literacy assessment

4.1. Research questions

This review aims to systematically examine existing instruments used to measure data literacy. By focusing on their structure, validation methods and contextual relevance, the review seeks to highlight best practices, identify limitations and uncover gaps that can inform future developments in this field. To explore the current state of measurement instruments for data literacy, the review addresses the following research questions:

RQ1: Which target groups are addressed by existing instruments for measuring data literacy?
RQ2: Which competencies are addressed in existing instruments for measuring data literacy?
RQ3: How have these instruments been validated in terms of reliability, validity, and contextual relevance?

4.2. Data sources and search strategy

To identify studies focusing on the measurement of data literacy, a systematic search was conducted on 2 April 2025 across eight academic databases. These databases were selected to cover a wide range of research areas relevant to education, interdisciplinary studies and digital competencies. The following databases were included: ERIC, EBSCOhost, Web of Science, Springer Link, Taylor & Francis Online, IEEE Xplore, and BASE. In addition, Google Scholar was used to complement the database search and to capture further academic and grey literature that might not be indexed elsewhere.

We have followed PRIMSA recommendations (Page et al., 2021) and a single author has conducted the search process. We have mitigated the potential problem of missing important studies in this stage by applying multiple search strategies (snowball search, AI based search).

The development of the search strategy followed an iterative process, informed by initial test searches and ongoing relevance checks. The educational context was not explicitly restricted, as it was often implicitly addressed within the studies, allowing for a broader and more inclusive scope.

Searches combining "data literacy" with terms like "assessment," "measurement," or "scale" in titles and keywords yielded relatively few but mostly relevant results. Sample-based relevance assessments showed that these terms consistently led to moderate to high relevance across databases. To expand coverage, additional terms were tested in abstract and full-text fields. Among these, "questionnaire" proved particularly effective, whereas broader terms like "test" and "survey" produced large volumes of irrelevant results and were therefore excluded.

The final strategy included studies with "data literacy" or "Datenkompetenz" in the title and at least one measurement-related term ("assessment", "measurement", "questionnaire", "scale" or "Messinstrument") in all searchable fields. This approach ensured a clear conceptual focus while still capturing studies that described relevant instruments in the methodology or results sections rather than in abstracts.

All search strings were adapted to the syntax and field structures of each database (see Appendix Table A. 2). No filters were applied for time period or publication type, although only studies in English or German were considered. Google Scholar was used as an additional source via focused title searches, due to its broad indexing and limited filtering options. Reference lists of included full-text articles were reviewed to identify further relevant studies. In addition, keyword alerts on selected platforms helped identify newly published studies during the review process.

4.3. Selection process

The selection of studies followed a structured, multi-phase procedure in line with the PRISMA 2020 guidelines. All records identified through database and supplementary searches were imported into a reference management system. After removing duplicates (n = 169), n = 315 unique records remained for initial screening.

In the first phase, titles and abstracts were screened against predefined inclusion and exclusion criteria. Records clearly unrelated to the measurement of data literacy or addressing other constructs were excluded, leaving n = 31 potentially eligible articles for full-text review.

During the subsequent full-text screening, studies were assessed in detail to determine whether they reported an instrument explicitly designed to assess data literacy and included at least basic evidence of validation. The applied criteria are summarized in Table 3. A total of 25 studies were excluded at this stage for not meeting these core requirements. This step also included a critical appraisal of conceptual relevance for studies that formally met the inclusion criteria but raised concerns regarding their alignment with the overarching aims of the review. While no fixed set of competencies was required, instruments were expected to reflect a comprehensive and transferable understanding of data literacy. Studies were excluded if they focused solely on isolated subdimensions or described highly context-specific, one-time interventions that lacked generalizability or theoretical depth. This additional review also addressed studies with unclear or ambiguous reporting of validation procedures, which required further discussion to determine whether they sufficiently met the minimum quality standards. As a result, only instruments with sufficient conceptual scope, generalizability and methodological soundness were retained for inclusion. In total, n = 6 studies were retained.

The overall screening process followed a structured procedure: initial title and abstract screening was conducted by a single reviewer, followed by full-text assessments. Throughout all phases, decisions involving uncertainty, particularly during critical appraisal, were discussed collaboratively with a second researcher to ensure consistent application of the criteria. The entire process, including rationales for inclusion and exclusion, was thoroughly documented.

To complement the database search, both backward and forward snowballing were conducted based on the included full-text articles, but this did not yield any additional eligible studies. An additional search using AI-based tools likewise did not identify further relevant records.

The application of the inclusion and exclusion criteria was discussed by both authors. Final selection of included and excluded articles were discussed in a consultation session with a second researcher to ensure transparency and consistency in the application of the predefined criteria.

An overview of the screening and selection process is provided in the PRISMA flow diagram (Fig. 3).

4.4. Results

4.4.1. Instruments and target groups

The six included studies address a range of target groups for the assessment of data literacy. Table 4 provides an overview of the application context, target group and type of instrument for each study.

Three instruments were developed specifically for the higher education context. Kirby and Lewis (2024) designed a self-report scale targeting university students as well as working professionals, emphasizing the need for data literacy in both academic and workplace settings. Qiao et al. (2024) focused on STEM students, developing a multiple-choice test grounded in scientific data use and

Table 3
Inclusion and exclusion criteria by category – assessment instruments.

Category	Inclusion Criteria	Exclusion Criteria
Instrument Scope	Instrument was specifically designed to assess data literacy as an explicitly named and central construct	Instrument assessed other constructs or broader literacies without clearly focusing on data literacy
Validation Requirements	Reported at least internal consistency as evidence of instrument reliability or validity; further validation types (test–retest reliability, construct and predictive validity) were documented where available	Did not report internal consistency or any other form of validation relevant to the instrument's quality or reliability
Documentation Quality	Included sufficient detail on the development, structure, or validation of the instrument (e.g., item design, validation methods, scoring procedures)	Lacked sufficient documentation regarding the instrument's design, structure, or validation process
Publication Type	Peer-reviewed journal articles, conference papers, book chapters, selected grey literature (e.g., institutional reports)	Commentary or opinion pieces and other non-research publications without original empirical or conceptual contribution
Time Scope	Any	-
Language/ Geographical Scope	Written in English or German, any geographical origin	Written in languages other than English or German

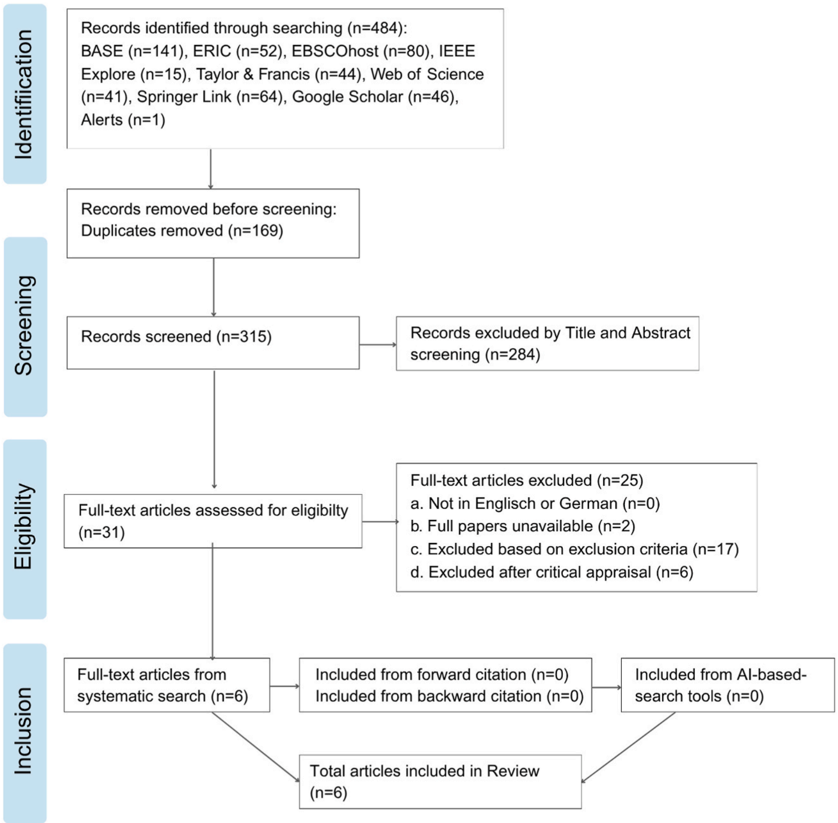

Fig. 3. PRISMA flow diagram of the literature selection process - assessment instruments.

Table 4
Application contexts, target groups and types of instruments in the reviewed studies.

	Application Context	Author (Year)	Target Group	Type of Instrument
1	Higher Education & Workforce	Kirby & Lewis (2024)	University students and working professionals	Self-report scale (Likert)
2		Kim et al. (2025)	University students (undergraduate and graduate)	Self-report scale (Likert)
3	University Students (Discipline-Specific)	Qiao et al. (2024)	University students in STEM disciplines	Multiple-choice test
4	Teacher Education	Donate-Beby et al. (2025)	Primary and secondary school teachers	Self-report scale (Likert)
5		Duygulu et al. (2025)	Teachers and principals (focus on institutional data literacy)	Self-report scale (Likert)
6	Public Sector Context	Ongena (2023)	Public sector professionals and administrative staff	Self-report scale (Likert)

Note. Application Context refers to the primary educational or professional setting for which the instrument was developed or validated. *Target Group* refers to the primary population for which the instrument was designed or evaluated as described by the original authors. Where applicable, additional intended user groups explicitly identified in the original publication are also reported.

domain-specific reasoning. Most recently, Kim et al. (2025) validated the Data Literacy Self-Efficacy Scale (DLSES) with undergraduate and graduate students at four-year institutions, aiming to assess self-perceived data literacy across a broad range of disciplines.

Two instruments were developed for the school education sector. Donate-Beby et al. (2025) introduced a self-report instrument tailored to primary and secondary school teachers, with a particular focus on ethical and pedagogical dimensions of data use. Duygulu et al. (2025) developed a scale targeting both teachers and school principals, emphasizing conceptual knowledge of data practices and data management processes in educational institutions.

Finally, Ongena (2023) developed a self-report scale for public sector professionals and administrative staff, acknowledging the increasing role of data-informed work practices in government and administration.

4.4.2. Addressed competencies

The analysis of the instruments revealed that they vary considerably in terms of the competency dimensions they address (Fig. 4). While no instrument covers the full range of identified competencies, several recurring patterns emerge.

The most frequently represented dimensions are *Identifying & Accessing Data, Data Preparation,* and *Data Visualizing*, each included in six instruments (100%). *Data Analysis* is addressed by five instruments (83%). *Ethics* and *Communication & Storytelling* is covered in four instruments (67%), while *Data Management & Organization* and *Evaluating Data* appear in three instruments (50%).

Other dimensions are addressed less frequently: *Understanding Data, Data Collection* and *Data-Driven Action* are each included in two instruments (33%). *Real-world-problem-solving context, Ask Questions from data, Data Sharing & Citation, Data Culture* and *Assessing Societal Implications* are each represented in only one instrument (17%).

The number of addressed dimensions per instrument ranges from six (Kirby & Lewis, 2024) to 11 (Qiao et al., 2024).

Duygulu et al. (2025) covers a narrower set of competencies, mainly focusing on *Understanding Data, Data Preparation,* and *Data Management & Organization*. The instrument is designed to assess conceptual knowledge and basic practices of data use in schools, with an emphasis on teachers' and school administrators' awareness of data-related terminology, storage tools and management processes.

Ongena (2023) addresses 10 dimensions, including both technical and reflective aspects such as *Ethics, Evaluating Data,* and *Assessing Societal Implications*. The instrument comprises several subdimensions of data literacy and was developed for professionals in the public sector.

Qiao et al. (2024) addresses the broadest range of dimensions among the instruments reviewed and is the only one using a multiple-choice format. It was developed for university students in STEM disciplines and focuses on declarative knowledge related to data use, including areas such as *Data Interpretation*, *Communication,* and *Data Sharing & Citation.*

4.4.3. Validation of the instruments

All six studies included in the review report on aspects of instrument validation, with a particular focus on internal consistency and construct validity (Table 5). Internal consistency was reported in all cases, typically through Cronbach's alpha coefficients exceeding the commonly accepted threshold of .70. This suggests that scale coherence and internal reliability were consistently addressed during instrument development.

While most instruments did not assess test–retest reliability, one study (Duygulu et al., 2025) included a test–retest procedure, reporting strong correlation coefficients across two measurement points. None of the instruments reported evidence of predictive validity, such as correlations with academic outcomes, behavioral indicators or external assessments.

Regarding construct validity, all instruments included some form of supporting evidence. While Donate-Beby et al. (2025) relied solely on exploratory factor analysis (EFA), all other studies conducted confirmatory factor analyses (CFA), with some providing additional validation indicators. Kim et al. (2025), for example, reported composite reliability and average variance extracted (AVE) in addition to EFA and CFA, based on a large sample of over 1800 university students. Duygulu et al. (2025) also combined EFA and CFA to examine the factorial structure of their scale. Kirby and Lewis (2024) and Qiao et al. (2024) applied item response theory to evaluate item characteristics and measurement precision. Ongena (2023) combined CFA of eleven first-order constructs with partial least squares structural equation modeling (PLS-SEM) to validate a formative second-order model.

Several studies linked item development to specific professional or educational contexts, using expert feedback (Qiao et al., 2024; Donate-Beby et al., 2025; Duygulu et al., 2025; Ongena, 2023), theoretical frameworks (Kim et al., 2025; Kirby & Lewis, 2024) or pilot testing (Qiao et al., 2024; Kim et al., 2025; Donate-Beby et al., 2025) to enhance content validity.

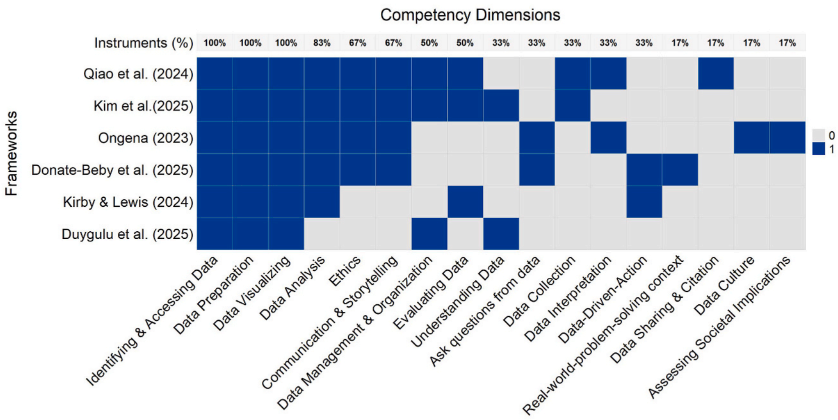

Fig. 4. Heatmap of competency dimensions covered across reviewed instruments.

Table 5
Validation types reported across reviewed data literacy instruments.

–	Author (Year)	Validation			
		Internal Consistency	Test-Retest Reliability	Construct Validity	Predictive Validity
1	Kirby & Lewis (2024)	✓	×	✓	×
2	Kim et al. (2025)	✓	×	✓	×
3	Qiao et al. (2024)	✓	×	✓	×
4	Donate-Beby et al. (2025)	✓	×	✓	×
5	Duygulu et al. (2025)	✓	✓	✓	×
6	Ongena (2023)	✓	×	✓	×

4.5. *Discussion of the findings*

4.5.1. *Addressed competencies*

The instruments show a clear focus on technical and procedural aspects of data literacy. Dimensions such as *Identifying & Accessing Data, Data Preparation* and *Data Visualizing* were included in nearly all instruments, while *Data Analysis* appeared in five out of six, highlighting their perceived importance for working with data. In contrast, dimensions related to critical reflection, ethical reasoning, and societal awareness were addressed much less frequently.

The uneven coverage of competencies may reflect differences in the intended application contexts and underlying goals of the instruments. In some cases, a focus on administrative structures or institutional processes (Duygulu et al., 2025) coincides with an emphasis on procedural and conceptual knowledge. Other instruments incorporate normative elements such as ethical considerations (Donate-Beby et al., 2025; Ongena, 2023), but there is no consistent pattern linking target groups or contexts to specific sets of competencies.

The results suggest that current instruments tend to emphasize selected aspects of data literacy, particularly those related to technical and procedural skills. While these areas are clearly relevant, other dimensions, such as data interpretation, ethical considerations and broader societal implications, receive comparatively little attention. This raises the question of whether existing instruments fully capture the multifaceted nature of data literacy across different contexts.

4.5.2. *Instrument validation*

All reviewed instruments reported internal consistency, typically in the form of Cronbach's alpha values. Most studies also included evidence of construct validity, often using confirmatory factor analysis (CFA). However, the range of validation strategies remained relatively narrow. Only one instrument (Duygulu et al., 2025) included a test–retest procedure to assess temporal stability, and no study reported validation efforts beyond these commonly applied approaches.

The relatively narrow range of validation strategies raises the question of whether the available evidence sufficiently supports the reviewed instruments. However, the relevance of different forms of reliability and validity evidence depends on an instrument's intended use and context of application.

Moreover, the diversity of analytical approaches (e.g., CFA, PLS-SEM, IRT, Rasch modeling) complicates direct comparisons between instruments. While some studies employed advanced techniques to examine latent structures or item characteristics, the use of these methods varied considerably. This variation may reflect differing conceptions of data literacy and the range of formats used to

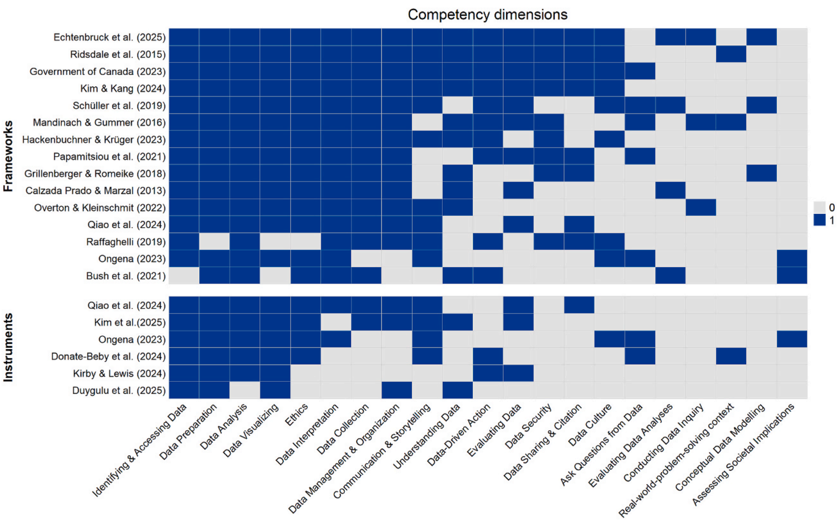

Fig. 5. Comparison of competency dimensions across data literacy frameworks and assessment instruments.

operationalize it, such as self-report scales and knowledge tests.

5. General discussion

This review set out to provide an integrated understanding of how data literacy is conceptualized and assessed. In response to ongoing ambiguities in the literature, the aim was to explore which dimensions of data literacy are emphasized across existing frameworks and how these are reflected in current assessment practices. By bringing together conceptual models and empirical instruments, the review offers a comprehensive perspective on the current state of research and identifies key challenges for both theoretical development and practical application.

5.1. Discussion of key findings

The comparative analysis of frameworks and assessment tools reveals both overlap and divergence in how data literacy is defined and assessed. Fig. 5 provides an integrated overview of the competency dimensions represented across frameworks and assessment instruments. This illustrates which competency dimensions are addressed by each framework and which are operationalized in the existing assessment instruments. Overall, the comparison reveals a shared set of competency dimensions across both frameworks and assessment instruments, while also highlighting important conceptual differences. Technical and procedural dimensions, such as *Data Analysis, Visualization, Preparation,* and *Identifying & Accessing Data*, are consistently emphasized, indicating a shared focus on data handling and processing across the data lifecycle.

Ethical aspects are prominently addressed in most frameworks and appear in four of the six analyzed instruments. This indicates a broad consensus on the importance of responsible data use as a key component of data literacy. However, the way ethics is incorporated differs: frameworks often present it as a cross-cutting concern embedded across dimensions, whereas instruments typically operationalize it as a separate subscale.

Beyond these commonly represented dimensions, substantial variation emerges. Frameworks tend to incorporate a broader conceptual landscape, including dimensions related to data interpretation, communication and action. Several models also integrate context-specific elements, such as *Real-World-Problem-Solving Context, Conceptual Data Modeling,* or *Data Culture*, which are largely absent from current measurement tools.

In contrast, assessment instruments adopt a more selective focus. While some tools, such as the instrument developed by Qiao et al. (2024), aim to capture a broad range of competencies, most instruments adopt a much narrower operationalization of data literacy. This may reflect their contextual orientation or their intended use within specific educational or professional settings. Critical and reflective dimensions, such as *Evaluating Data, Evaluating Data Analyses* and *Assessing Societal Implications* or even *Data Interpretation* are either rarely addressed or entirely absent from current assessment tools. This indicates a broader imbalance in how data literacy is conceptualized and assessed.

The empirical evidence supporting current instruments also remains relatively limited in scope. Although standard psychometric procedures such as internal consistency (Cronbach's alpha) and confirmatory factor analysis are frequently applied, other forms of validation remain rare. Only one instrument has been tested for temporal stability, and none has been examined in terms of predictive validity. None of the instruments have provided details on measurement invariance. While these findings provide a general impression on the spectrum of validation approaches, this should not be a judgement of individual studies, whose approach depends on the instrument's measurement model, developmental stage, intended population, application context, and the interpretations or decisions based on its scores. Accordingly, the absence of a particular type of evidence indicates an area that has not been examined or reported, but does not necessarily imply inadequate instrument quality.

Moreover, while some studies reference theoretical frameworks, the connection between the tested items and the broader construct of data literacy is often weak or unclear. As a result, current instruments do not only capture a reduced set of competencies but are also supported by a relatively narrow range of empirical evidence, limiting conclusions about their suitability across different contexts and applications.

Beyond these validation gaps, the assessment methods themselves could also present certain limitations. While self-report formats dominate the few existing objective test formats offer only limited insight into whether learners can apply data literacy skills in meaningful, real-world situations. Consequently, existing instruments may only partially reflect the situated and action-oriented nature of data literacy, which several frameworks conceptualize as central.

Overall, the results indicate a shared set of core dimensions across frameworks and tools, primarily centered on technical skills and ethical awareness. At the same time, frameworks more clearly reflect the complexity and contextuality of data literacy by incorporating reflective, communicative and civic elements that remain underrepresented in current assessment practices. While frameworks present a broad conceptual landscape of data literacy, some dimensions nevertheless appear only sporadically across models. This raises the question of which dimensions are truly essential for developing data literacy and whether the observed conceptual breadth reflects genuine educational needs or rather a lack of consensus regarding core competencies.

5.2. Conceptual overlaps with adjacent literacies

A closer examination of the identified competency dimensions reveals significant conceptual overlaps between data literacy and other established literacies, most notably information literacy and statistical literacy.

Information literacy is commonly defined as the ability to recognize when information is needed and to locate, evaluate and use it effectively and ethically (American Library Association [ALA], 1989). The ACRL Standards (2000) elaborate this into five core competencies: defining an information need, accessing information efficiently, evaluating sources critically, using information purposefully and understanding its ethical, legal and social dimensions. In the reviewed frameworks and measurement instruments, clear overlaps with these principles emerge. Dimensions such as *Identifying & Accessing Data, Evaluating Data, Ethics,* and *Data Management & Organization* reflect these competencies. Locating relevant datasets and assessing their quality parallels the processes of information access and evaluation. Likewise, responsible handling of data, including attribution, privacy, documentation, and data sharing, extends traditional information literacy into data-specific contexts, as reflected in dimensions such as *Data Management & Organization* and *Data Sharing & Citation*. This close alignment is also reflected in Carlson et al. (2011), who describe data information literacy as a natural extension of information literacy, emphasizing the need to equip learners with the skills to manage, document and ethically share data within increasingly data-intensive research environments.

Statistical literacy generally refers to the ability to understand, interpret, and critically evaluate statistical information, including uncertainty, variation and the validity of claims based on data (Gal, 2002; Schield, 2005). This perspective aligns closely with several technical dimensions identified in data literacy frameworks and measurement instruments. In particular, *Data Preparation, Data Analysis, Data Interpretation* and *Data Visualizing* emerged as core components present in nearly all reviewed models. These dimensions reflect the central concerns of statistical reasoning and suggest that statistical literacy forms a foundational layer of data literacy. As Gould (2017) argues, the boundaries between the two are increasingly blurred. He proposes an expanded view of statistical literacy that includes aspects such as data privacy, the traceability of data origins and algorithmic reasoning, ultimately suggesting that "data literacy is statistical literacy" in a data-driven society.

Beyond the overlaps with information and statistical literacy, several dimensions identified in the reviewed frameworks suggest that data literacy encompasses a broader and more distinct set of competencies. Components like *Communication & Storytelling,* and *Data-Driven Action* underscore the communicative and actionable nature of data literacy. They point to a shift from passive understanding toward active transformation, enabling individuals not only to make sense of data, but to use it for informed decision-making and purposeful intervention.

In addition, categories such as *Conducting Data Inquiry, Ask Questions from Data,* and *Real-world-problem-solving context*, coded only infrequently, could raise the question of whether they need to be included as distinct competencies. Together, however, they reflect an inquiry-oriented approach to data literacy. This approach is characterized by the capacity to formulate data-based questions, apply data skills to real-world problems, and use data as a tool for exploration and discovery. These characteristics arguably represent a distinctive feature of data literacy.

By contrast, certain additional dimensions that appeared only sporadically in the reviewed frameworks, like *Assessing Societal Implications* and *Data Culture*, raise concerns about their status within data literacy and lead to further conceptual questions. While the former reflects a growing awareness of the broader consequences of data use, it is unclear to what extent this reflective stance is considered a core component of data literacy. *Data Culture,* as defined here, encompasses the ability to recognize the value of data and to contribute to an environment that supports responsible, reflective, and strategic use of data, often within organizational settings, which blurs the boundaries between personal competencies and broader contextual factors. These ambiguities point to the evolving and contested nature of the construct, reinforcing the need for further clarification.

Taken together, the findings illustrate that data literacy draws heavily on competencies traditionally associated with information and statistical literacy. However, the review also reveals additional dimensions, such as *Conducting Inquiry*, *Communication* and *Data-driven Action*, that are not commonly addressed in those literacies and reflect the applied, problem-oriented character of data literacy. These distinctive elements, if considered as essential, suggest that data literacy cannot be fully captured as a mere combination of existing literacies. Rather, it should be regarded as a distinct and integrative competence, provided that its core dimensions are clearly defined and conceptually delimited.

5.3. Proposed conceptual framework: core-periphery structure of data literacy (COPERDALE)

Based on the findings of the preceding systematic reviews, we propose a conceptual framework that synthesizes the recurring dimensions of data literacy identified across existing frameworks and measurement instruments into a three-level core-periphery structure. The COPERDALE framework intends to overcome issues identified in the analysis of existing frameworks by following Blömeke, Gustafsson, and Shavelson's continuum model of competence (Blömeke, Gustafsson, & Shavelson, 2015). Accordingly, we do not treat data literacy as either an individual attribute or an observable practice. Rather, we conceptualize it as a layered construct in which individual knowledge, skills, and dispositions form the resource base for situation-specific enactment, which in turn becomes meaningful and consequential only within particular socio-cultural, institutional, and ethical contexts.

The core comprises relatively context-independent domain-specific knowledge, skills, and abilities required to understand, acquire, prepare, analyze, interpret, visualize, communicate, and manage data. The second layer captures the mobilization of these resources in authentic data practices, which we summarize as conducting data inquiry. The outer layer reflects the socio-material, cultural, institutional, and ethical conditions under which data practices become meaningful, legitimate, and consequential (Pangrazio & Sefton-Green, 2020).

The proposed three-level structure builds on the understanding of competencies as context-dependent ability constructs whose successful application becomes evident only in authentic situations and through interaction with the environment (Klieme et al., 2008). Accordingly, the differentiation between the core, the inner periphery, and the outer periphery reflects increasing dependence on authentic situations as well as organizational and societal contexts.

This organization highlights that the three levels differ not only in their content but also in the extent to which their successful enactment depends on contextual factors. Consequently, educational formats can differentiate clearly between activities related to individual knowledge and skill acquisition, the application of these competencies within authentic inquiry contexts, and their relation to the societal, legal, and organizational contexts in which they are applied.

An overview of the proposed framework is presented in Fig. 6.

5.3.1. Core: fundamental skills

The core encompasses fundamental skills that are closest to the individual and can be acquired independently of context, allowing them to be applied across different scenarios.

These include, first, a conceptual understanding of data and its structures (Understanding Data), covering knowledge such as how data are generated, the different types, formats, and structures they can take or the potential value of data across different fields of application.

The targeted acquisition and evaluation of data (Data Acquisition) combines the categories Identifying & Accessing Data and Data Collection and also involves the critical assessment of data quality, origin, and suitability (Evaluating Data) to identify potential biases and methodological weaknesses.

Data Preparation involves cleaning, standardizing, transforming and merging raw data in preparation for Data Analysis, in which the data is evaluated in a structured manner using appropriate methods.

Data Interpretation refers to the critical classification and evaluation of analysis results in light of the specific application context, including, where relevant, the derivation of well-founded options for action. The implementation and monitoring of these measures (Data-Driven Action) are located on the outer periphery, as they are dependent more heavily on institutional structures or a broader societal context.

Data Visualization involves selecting appropriate forms of visualization and designing them in a clear and consistent manner to effectively convey structures, patterns and results in an understandable way.

Data Communication refers to presenting analysis results clearly, precisely and in a manner appropriate to the audience, including the selection of suitable forms of presentation and using relevant technical terminology correctly.

Data Management & Organization involves the systematic organization, documentation, backup and access control of data, including structured storage, versioning, metadata management, and measures to ensure integrity and long-term usability.

Data Ethics represents the responsible and thoughtful handling of data, ensuring protection, security, transparency and fairness in

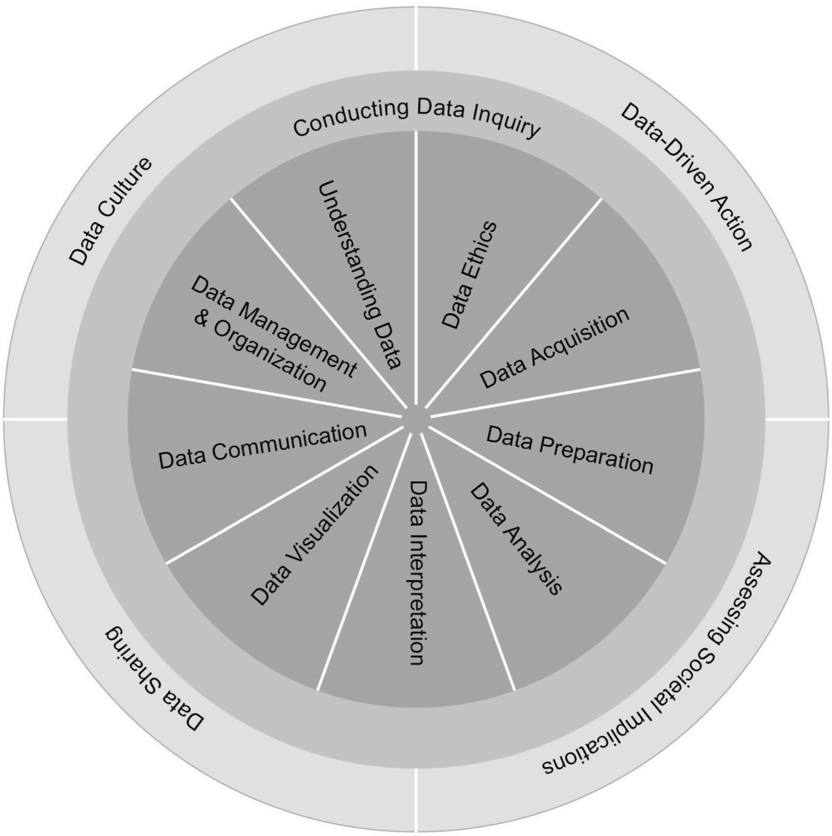

Fig. 6. Coperdale framework: Core–periphery structure of data literacy dimensions.

compliance with legal and professional standards (see Appendix Table C. 1 for a detailed overview of all dimensions and competencies).

5.3.2. Inner periphery: transfer and process level

The inner periphery represents the process level of the model, integrating the capabilities located in the core into a coherent data-based investigation process. This inquiry process describes the path from problem formulation to well-founded conclusions based on data (cf. Wild & Pfannkuch, 1999) and is summarized under the term Conducting Data Inquiry. It begins with the development of clear and relevant questions (Asking Questions from Data) and covers the essential steps from planning and analysis to the targeted communication of results. This level illustrates how the core competencies interlock and are applied effectively in real problem and action contexts. On this level, other knowledge and skills might have an influence on the success of the enactment of data literacy.

5.3.3. Outer periphery: systemic and context-dependent dimensions

The outer periphery encompasses competencies that extend beyond individual practice and are closely linked to organizational cultures, societal priorities, and governance structures. The dimensions located at this level are sometimes highly context-dependent and can only be understood to a limited extent as purely individual skills. They include Data Culture, promoting an environment that supports responsible, reflective and strategic use of data; Data-Driven Action, involving the initiation, implementation and review of data-based measures in coordination with organizational processes and stakeholders; Assessing Societal Implications, referring to the critical evaluation of the social, cultural and organizational impacts of data use.

The outer periphery also comprises Data Sharing, which entails the legal, ethical, and professional sharing of data while considering organizational and technical conditions. Although positioned here due to its strong dependence on institutional frameworks, several preparatory and ethical aspects are already addressed in the core, particularly within Data Management & Organization and Data Ethics.

5.3.4. Flexibility and measurability

The model combines a stable, context-independent core of fundamental skills with scope for context-specific extensions, particularly at the level of the outer periphery. This structure allows adaptation to different institutional, disciplinary, or societal conditions without altering the core competencies. The distinction between a stable core and context-dependent peripheries also clarifies which dimensions can be regarded as individual skills and which are shaped more strongly by institutional, social or political factors. While core competencies relate directly to an individual's abilities, some peripheral dimensions, especially those in the outer periphery, can only be measured to a limited extent at the individual level due to their dependence on specific contexts.

5.4. Implications for research and practice

The findings of this review point to several implications for both future research and practical applications in educational contexts. One central issue is the persistent misalignment between the conceptual breadth of data literacy frameworks and the narrower scope of current assessment instruments. Frameworks often aim to offer comprehensive, ideal-type models, whereas measurement instruments tend to pragmatically reduce data literacy to a set of measurable, frequently technical skills. As a result, current assessment practices may fall short of capturing the full scope of data literacy as envisioned in broader conceptual models. The proposed COPERDALE Framework addresses this gap by distinguishing between a stable, context-independent core of fundamental skills and context-dependent peripheries. This structure clarifies which competencies are most suitable for universal inclusion in assessments and which require adaptation to specific institutional or societal contexts.

In this context, future research should also reconsider how data literacy is assessed in practice. The dominance of self-report formats raises the question of whether such instruments can adequately capture the competencies they aim to measure, particularly when these involve applied skills rather than self-perceived abilities. Complementary approaches, such as performance-based tasks or objective test items, may provide a more accurate indication of whether learners can apply data literacy skills effectively. Developing such instruments, however, requires not only methodological innovation but also conceptual clarity about which aspects of data literacy are being targeted.

More broadly, the lack of consensus regarding core dimensions of data literacy continues to pose a major challenge for the field. As long as definitions remain diffuse and frameworks vary widely in their conceptual focus, efforts to design valid and meaningful assessments will be constrained. Advancing the field thus requires a more coordinated effort to clarify which competencies are essential and how they relate to one another within a coherent model of data literacy. The proposed framework offers one possible foundation for such coordination, while recognizing the need for empirical validation and potential adaptation across diverse contexts.

5.5. Limitations of this study

While this review offers a comprehensive synthesis of current frameworks and assessment approaches, certain limitations must be considered. First, the analysis was restricted to literature published in English and German. Although these languages cover a substantial share of the relevant discourse, publications in other languages may provide additional perspectives that were not captured. Second, data literacy is a rapidly evolving field. Despite setting up automated alerts across multiple platforms to monitor new developments, it is possible that some very recent studies were not included in the analysis. Nonetheless, the combination of a broad search strategy, the inclusion of grey literature, and continuous monitoring of new publications enhances the robustness of the findings and mitigates some of the limitations outlined above. A potential publication bias could have been existing in the published literature on data-literacy. While this problem is accounted for in a meta-analysis, there is no equivalent solution in a systematic literature review. Furthermore, the majority of frameworks and instruments stems from the higher education context while the adult education context is underrepresented in the set of identified frameworks and measurement instruments. The findings should also be interpreted in light of the conceptual boundaries applied during study selection. A broader conceptualization of eligible frameworks and assessment instruments might have included additional perspectives on data literacy.

6. Conclusion

The current definitional and theoretical fragmentation of frameworks and instruments for data literacy has been the motivation to conduct the two systematic literature reviews reported in this study. This review synthesized evidence from two systematic analyses to provide an integrated understanding of how data literacy is conceptualized and assessed. The findings show that while existing frameworks capture a broad range of competencies, assessment instruments typically focus on a narrower subset, often emphasizing technical skills such as data handling, analysis and visualization. Reflective and ethical dimensions, although present in many frameworks, are less frequently operationalized in measurement tools. Moreover, the dominance of self-report instruments and limited validation approaches suggest that existing tools may not yet capture the complexity of data literacy. While a lack of frameworks with sufficient coverage of different dimensions of data literacy has been identified in the article, the study has proposed an integrated framework with broader coverage as synthesis of the findings. This framework is based on the understanding of competence as a continuum between acquirable (latent) knowledge and skills, the application of those in a specific problem situation (data inquiry) and the social, ethical, legal and societal context in which those skills are developed and applied. The proposed Core–Periphery Framework offers a structured way to distinguish a stable, context independent core of skills from more variable peripheral dimensions. This distinction can help prioritize what to measure universally and what to tailor to institutional or disciplinary settings. Furthermore, it enables learning designers in higher and adult education to develop more tailored educational formats with more specific learning outcomes.

The proposed framework offers a theory-informed synthesis of dimensions of data-literacy, their relations and context-related layers. Future work in extension of the framework can one the one hand focus on defining complexity levels for the different dimensions and on the other hand conduct research on a logic of knowledge and skill-acquisition and a concept of sequencing.

The article has also identified a limited set of measurement instruments with a.) sufficient conceptual and theoretical coverage and b.) broader evidence of consistency, validity and reliability. Future work should focus on the development of measurement instruments based on the proposed framework and with a high level of validation quality. In general, future efforts should aim to strengthen the connection between conceptual models and assessment practices, expand the range of competencies that are operationalized in measurement tools and diversify methodological approaches beyond self-report instruments. This includes developing performance-based formats or objective test items, broadening validation procedures and ensuring that both technical and reflective dimensions of data literacy are adequately represented. The three-level structure of the framework can also serve as a basis to develop sector-specific competence models which share a joint knowledge and skill-portfolio (the core) but can be differentiated in the periphery parts of the framework.

Author contributions

Marlene Ganz: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing.

Marco Kalz: Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL Write in order to improve grammar and language. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors disclose that they do not have any conflict of interest to declare.

Appendix A. Database Search Syntax

Table A. 1
Search Strategy Syntax – Framework Review

Database	Search date	Syntax	Records
BASE	02 Feb 2025	tit:("data literacy" OR "Datenkompetenz") (Framework OR Kompetenzrahmen OR "competence model" OR Kompetenzmodell) doctype:(11* 12* 13 14 15 16 18* 1A)	148
ERIC	02 Feb 2025	title:("data literacy" OR "Datenkompetenz") AND (framework OR Kompetenzrahmen OR "competence model" OR Kompetenzmodell)title:("data literacy" OR "Datenkompetenz") AND (framework OR Kompetenzrahmen OR "competence model" OR Kompetenzmodell)	36
EBSCO Host	02 Feb 2025	TI ("data literacy" OR Datenkompetenz) AND (framework OR Kompetenzrahmen "competence model" OR Kompetenzmodell)	37
IEEE Xplore	02 Feb 2025	("Document Title":"data literacy" OR "Document Title":Datenkompetenz) AND ("All Metadata":framework OR "All Metadata":Kompetenzrahmen OR "All Metadata":"competence model" OR "All Metadata":Kompetenzmodell)	6
Taylor & Francies	02 Feb 2025	Title:("data literacy" OR Datenkompetenz) AND Anywhere:(framework OR Kompetenzrahmen OR "competence model" OR Kompetenzmodell)	37
Web of Science	07 Feb 2025	"data literacy" OR "Datenkompetenz" (Title) AND framework OR Kompetenzrahmen OR "competence model" OR Kompetenzmodell (All Fileds)	31
Google Scholar	02 Feb 2025	(intitle:"data literacy" OR intitle:Datenkompetenz) AND (intitle:framework OR intitle:Kompetenzrahmen OR intitle:"competence model" OR intitle:Kompetenzmodell)	68

Table A. 2
Search Strategy Syntax – Assessment Instruments

Database	Search date	Syntax	Records
BASE	02 Apr 2025	tit:("data literacy" OR "Datenkompetenz") AND (assessment OR measurement OR questionnaire OR scale OR Messinstrument) doctype:(11* 12* 13 14 15 16 18* 1A)	141
ERIC	02 Apr 2025	title:("data literacy" OR "Datenkompetenz") AND (assessment OR measurement OR questionnaire OR scale OR Messinstrument)	52
EBSCO Host	02 Apr 2025	TI ("data literacy" OR "Datenkompetenz") AND (assessment OR measurement OR questionnaire OR scale OR Messinstrument)	80
IEEE Xplore	02 Apr 2025	("Document Title":"data literacy" OR "Document Title":Datenkompetenz) AND ("All Metadata":assessment OR "All Metadata":measurement OR "All Metadata":questionnaire OR "All Metadata":scale OR "All Metadata": Messinstrument)	15
Taylor & Francies	02 Apr 2025	Title: ("data literacy" OR "Datenkompetenz") AND Anywhere: (assessment OR measurement OR questionnaire OR scale OR Messinstrument)	44
Web of Science	03 Apr 2025	"data literacy" OR Datenkompetenz (Title) AND assessment OR measurement OR questionnaire OR scale OR Messinstrument (All Fileds)	41
Google Scholar	02 Apr 2025	allintitle:"data literacy" OR Datenkompetenz AND assessment OR measurement OR questionnaire OR scale OR Messinstrument	46

Appendix B. Descriptive Overview of Included Data Literacy Frameworks

Table B. 1
Descriptive Overview of Included Data Literacy Frameworks

	Author (Year)	Official Framework Name	Purpose	Structure	Proficiency Levels
1	Bush et al. (2021)	-	To promote data literacy competencies for employability and citizenship.	Three competence areas (Reading & Understanding, Reflection, and Creation/ Design), each described across three requirement levels.	Reproduce; Connect relationships; Transfer
2	Calzada Prado & Marzal (2013)	-	To provide a reference framework for integrating data literacy into library information literacy programs.	Five core competencies, subdivided into eight competency areas with corresponding instructional contents.	-
3	Ridsdale et al. (2015)	-	To define a cross-sectoral data literacy competency framework for students, researchers, professionals, and policymakers.	Competency matrix with five key ability/ knowledge areas, competencies, and knowledge/tasks.	Conceptual; Core; Advanced

(continued on next page)

Table B. 1 (*continued*)

	Author (Year)	Official Framework Name	Purpose	Structure	Proficiency Levels
4	Schüller et al. (2019)	-	To support higher education institutions in fostering data literacy competencies.	Process-oriented framework comprising competence areas, each described by knowledge, skills, and attitudes.	Three competence levels (1–3)
5	Echtenbruck et al. (2025)	Data Literacy Initiative (DaLI)	To support the strategic integration of data literacy into higher education curricula.	Seven competence areas aligned with the research data life cycle, each comprising multiple competencies with associated knowledge and tasks/skills.	-
6	Qiao et al. (2024)	-	To develop a conceptual framework for science data literacy and support STEM curriculum development.	Three-dimensional competency framework comprising 11 elements.	-
7	Hackenbuchner & Krüger (2023)	DataLitˆMT Framework	To develop a data literacy competency framework for translation and specialized communication education.	Five-dimensional competency framework comprising multiple subdimensions.	Basic; Advanced
8	Overton & Kleinschmit (2022)	Data Science Literacy Framework (DSLF)	To integrate data literacy into public administration education.	Four-domain framework comprising one task-oriented domain (Data Task) and three support-oriented domains (Computation, Statistical, and Application and Systems Integration Knowledge).	-
9	Mandinach & Gummer (2016)	Data Literacy for Teachers (DLFT)	To define the knowledge, skills, and dispositions teachers need for effective and responsible data use.	Seven knowledge areas integrated with a five-component inquiry cycle for data use.	-
10	Raffaghelli (2019)	-	To conceptualize data literacy as a core dimension of educators' professional competence.	Six competency dimensions adapted from the DigCompEdu framework.	A1 Newcomer; A2 Explorer; B1 Integrator; B2 Expert; C1 Leader; C2 Pioneer
11	Papamitsiou et al. (2021)	-	To support the development of educational data literacy competences for instructional designers and e-tutors.	Six dimensions (one generic ethical/legal and five data-related) comprising 21 competence statements, organised as an inquiry cycle.	-
12	Grillenberger & Romeike (2018)	-	To define data literacy competencies and support their integration into school computer science education.	Two-dimensional competency matrix (4 × 4) comprising four content areas and four process areas.	-
13	Kim & Kang (2024)	-	To guide data literacy development and inform educational policy and curricula.	Five categories comprising 23 competencies with associated sub-components.	-
14	Ongena (2023)	-	To support responsible and effective data use in the public sector.	Five higher-order dimensions comprising 11 first-order skills.	-
15	Government of Canada (2023)	-	To guide data competency development across the public service workforce.	Four categories comprising 13 competencies with associated indicators.	Foundational; Intermediate; Advanced

Appendix C. Overview of the COPERDALE Framework

Table C. 1
Dimensions and Competencies of the COPERDALE Framework

Core: Fundamental Skills	Description & Competencies	
Understanding Data	Fundamental conceptual understanding of data.	• Understanding Data Generation: Knowing how data is generated. • Recognizing Data Types and Structures: Understanding different data types and structures. • Understanding Data Formats and Metadata: Knowing common data formats and the role of metadata for interpretation and reuse. • Interpreting Measurement Levels: Distinguishing between different types of scales and understanding their implications for analysis. • Awareness of Data Quality and Bias: Knowing basic criteria of data quality and developing awareness of possible biases. • Recognizing the Value of Data: Assessing the potential value and usefulness of data in various contexts
Data Acquisition	Acquiring data by identifying and using existing sources or by collecting new data.	• Identifying Data Sources: Researching and selecting relevant internal or external data sources.

(continued on next page)

Table C. 1 (*continued*)

Core: Fundamental Skills	Description & Competencies	
Data Preparation	Preparing raw data for analysis.	• Accessing Data: Accessing data while considering technical interfaces, licensing conditions and access rights • Collecting Data: Gathering data using appropriate methods and standards. • Evaluating Data Quality: Critically assessing the quality, provenance and suitability of data, including completeness, timeliness and consistency, as well as recognizing potential biases, methodological weaknesses or limitations and evaluating their impact. • Cleaning Data: Detecting and correcting errors, inconsistencies, duplicates, missing values and outliers. • Standardizing Data: Harmonizing formats, units, naming conventions and coding schemes. • Transforming Data: Aggregating, normalizing, encoding or reshaping data for analytical purposes. • Integrating Data Sources: Combining datasets from different sources while managing join conditions, overlaps and conflicts. • Documenting Data Preparation: Recording applied transformations and cleaning procedures to ensure transparency and reproducibility.
Data Analysis	Systematic evaluation of data, ranging from structured exploration to the presentation of robust results.	• Exploratory Data Analysis: Visually and statistically examining data to identify patterns, trends and anomalies • Selecting Analytical Methods: Choosing appropriate statistical, algorithmic or model-based approaches • Applying Analytical Techniques: Conducting analyses with the chosen methods and interpreting the results correctly • Assessing Assumptions and Uncertainties: Evaluating key assumptions and potential uncertainties of the analysis
Data Interpretation	Placing and evaluating analysis results within the respective analytical or research context.	• Assessing Reliability and Validity: Checking whether results are consistent and methodologically solid • Recognizing Limitations and Bias: Identifying limitations of the data or methodology and evaluating their impact • Exploring Alternative Interpretations: Considering different plausible interpretations of the results • Relating Findings to Existing Knowledge: Linking results to existing knowledge or theoretical models • Deriving Actionable Recommendations: Translating findings into practical recommendations where appropriate
Data Visualizing	Designing clear and appropriate representations of data to effectively communicate structures, patterns and results.	• Selecting Appropriate Visualization Types: Choosing suitable chart types or graphical representations for the given data structure • Applying Effective Visual Encoding: Using colors, shapes, sizes or positions effectively to clearly convey information • Ensuring Consistency and Clarity: Designing axes, scales and labels consistently and understandably
Data Communication	Ability to communicate analysis results clearly, precisely and in a way that is appropriate for the target audience.	• Summarizing Key Findings: Presenting key results in a concise and accessible manner. • Adapting to Audience: Tailoring the level of detail, terminology and format to the needs of different audiences. • Using Appropriate Media and Formats: Employing written reports, presentations, dashboards or verbal communication effectively. • Supporting Communication with Evidence: Backing up statements with data.
Data Management & Organization	Systematic organization, documentation, safeguarding, and access control of data.	• Structuring Data Storage: Systematically storing and organizing data • Version Control: Transparently documenting changes to datasets • Metadata Creation: Providing contextual information to ensure find ability and reusability • Data Security Measures: Preventing unauthorized access • Archiving & Retention: Ensuring long-term preservation and legally compliant retention of data
Ethics	Responsible and ethically informed handling of data.	• Protecting Privacy: Safeguarding data privacy and ensuring confidentiality • Ensuring Fairness: Identifying, preventing and addressing biases • Maintaining Transparency: Disclosing methods, data sources and assumptions • Adhering to Legal Standards: Complying with relevant legal requirements
Inner periphery: Transfer and process level	Application and integration of core competencies within a structured process of data-driven investigation. This process begins with formulating a relevant research question based on a real-world or practice-oriented problem and continues through planning and conducting data collection, analyzing, interpreting and visualizing data, as well as communicating the results. The process level serves as a bridge between isolated individual competencies and their coherent application in a specific context. The focus is not on mastering single skills in isolation, but on their combined use to provide well-founded answers to data-related questions.	

(*continued on next page*)

Table C. 1 (*continued*)

Core: Fundamental Skills	Description & Competencies
Outer periphery: Systemic and context-dependent dimensions	
Data Culture	Contribution to an environment that fosters responsible, reflective, and strategic engagement with data. Includes shared standards and routines, clear role definitions, opportunities for learning and feedback, and structures that support openness, quality assurance, and continuous improvement.
Data-Driven Action	Ability to initiate, implement, monitor, and critically evaluate actions based on data. On an individual level, this involves deriving and assessing possible courses of action. In collaborative and organizational contexts, it includes aligning with stakeholders, integrating into strategic processes, and using shared resources. The dimension also encompasses mechanisms for continuous adaptation and improvement based on new data.
Assessing Societal Implications	Ability to critically assess the potential societal, cultural and organizational impacts of data use and data-driven decisions. This includes analyzing possible positive and negative effects for individuals, communities, and institutions, recognizing unintended consequences and considering ethical, legal and social frameworks when developing data-driven processes and structures.
Data Sharing	Ability to share data within existing organizational and technical structures in a legal, ethical and professional manner. This includes deciding which data can be released based on defined policies and licenses, selecting appropriate platforms and formats and documenting data in a way that ensures safe and transparent reuse by others. Implementation requires both individual skills in documentation and citation as well as integration into processes that ensure data quality, security and long-term reusability.

Data availability

No data was used for the research described in the article.

References

American Library Association. (1989). Presidential committee on information literacy: Final report. Retrieved from https://www.ala.org/acrl/publications/whitepapers/presidential. (Accessed 21 July 2025).

Association of College & Research Libraries, American Library Association. (2000). *Information literacy competency standards for higher education*. ACRL. Retrieved from https://alair.ala.org/items/294803b6-2521-4a96-a044-96976239e3fb. (Accessed 21 July 2025).

Bhargava, R., & D'Ignazio, C. (2015). Designing tools and activities for data literacy learners. In *The data literacy Workshop at WebScience 2015*. Oxford, UK.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. Competence viewed as a Continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bush, A., de Gruisbourne, B., Matzner, T., & Schulz, C. (2021). Data literacy: Kompetenzrahmen für Hochschulen. Retrieved from https://www.campus-owl.eu/fileadmin/campus-owl/dalis/documents/Kompetenzrahmen_Workingpaper210921.pdf. (Accessed 3 March 2025).

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

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.

Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., & Cutumisu, M. (2023). Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice, 30*(1), 76–96.

Deahl, E. (2014). Better the data you know: Developing youth data literacy in schools and informal learning environments. *SSRN*. https://doi.org/10.2139/ssrn.2445621 [Preprint].

Donate-Beby, B., García-Peñalvo, F. J., Amo-Filva, D., & Aguayo-Mauri, S. (2025). Filling the gap in K-12 data literacy competence assessment: Design and initial validation of a questionnaire. *Computers in Human Behavior Reports, 17*, Article 100583. https://doi.org/10.1016/j.chbr.2024.100583

Duygulu, A., Doğan, S., & Yıldız, S. (2025). Data literacy at school: A scale development study. *Journal of Theoretical Educational Science, 18*(1), 106–130. https://doi.org/10.30831/akukeg.1472451

Echtenbruck, M. M., Fühles-Ubach, S., Naujoks, B., & Kaliva, E. (2025). A data literacy competence model for higher education and research [Preprint]. *arXiv*. https://arxiv.org/abs/2504.15690.

Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia, 13*(1), 93–106.

Fleischer, J., Wirth, J., Rumann, S., & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz: Analyse von Aufgabenprofilen. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung: Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (Zeitschrift für Pädagogik, 56* pp. 239–248). Weinheim/Basel: Beltz. https://doi.org/10.25656/01:3432. Beiheft.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., & Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship, 28*(2), 112–127. https://doi.org/10.1080/08963568.2023.2171552

Glaser, R., Chudowsky, N., & Pellegrino, J. W. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press. https://doi.org/10.17226/10019.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22–25. Retrieved from https://iase-web.org/documents/SERJ/SERJ16(1)_Gould.pdf. (Accessed 21 July 2025).

Government of Canada. (2023). *Data competency framework (DDN3-J03)*. Canada School of Public Service. Retrieved from https://www.csps-efpc.gc.ca/tools/jobaids/data-competency-framework-eng.aspx. (Accessed 17 August 2025).

Grillenberger, A., & Romeike, R. (2018). Developing a theoretically founded data literacy competency model. In *Proceedings of the 13th workshop in primary and secondary computing education* (pp. 1–10).

Hackenbuchner, J., & Krüger, R. (2023). DataLitMT – Teaching data literacy in the context of machine translation literacy. In *Proceedings of the 24th annual conference of the european association for machine translation* (pp. 285–293). Tampere, Finland: European Association for Machine Translation. Retrieved from https://aclanthology.org/2023.eamt-1.28. (Accessed 17 August 2025).

Kim, J., Hong, L., Evans, S., & Yoon, A. (2025). University students' self-assessment of data literacy: A validation study. *PLoS One, 20*(4), Article e0322104. https://doi.org/10.1371/journal.pone.0322104

Kim, S., & Kang, H. (2024). Proposed data literacy competency framework through literature analysis. *International Journal of Knowledge Content Development & Technology, 14*(3), 115–140. https://doi.org/10.5865/IJKCT.2024.14.3.115
Kirby, J., & Lewis, J. A. (2024). Developing a brief data literacy measure using CFA and IRT. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006241300738
Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik, 56*(6), 10–13.
Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Hogrefe.
Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik, 52*(6), 876–903.
Krumm, S., Mertin, I., & Dries, C. (2012). *Kompetenzmodelle (Praxis der Personalpsychologie, Band 27), KG*. Göttingen: Hogrefe Verlag GmbH & Co.
Lee, J., Alonzo, D., Beswick, K., Abril, J. M. V., Chew, A. W., & Oo, C. Z. (2024). Dimensions of teachers' data literacy: A systematic review of literature from 1990 to 2021. *Educational Assessment, Evaluation and Accountability, 36*(2), 145–200. https://doi.org/10.1007/s11092-024-09435-8
Livingstone, S. (2004). The challenge of changing audiences: Or, what is the audience researcher to do in the age of the internet? *European Journal of Communication, 19*(1), 75–86. https://doi.org/10.1177/0267323104040695
Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37. https://doi.org/10.3102/0013189X12459803
Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007
Niegemann, H. M., Domagk, S., Hessel, S., Hein, A., Hupfer, M., & Zobel, A. (2008). *Kompendium multimediales Lernen*. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-540-37226-4. X.media.press.
Ongena, G. (2023). Data literacy for improving governmental performance: A competence-based approach and multidimensional operationalization. *Digital Business, 3*(1), Article 100050. https://doi.org/10.1016/j.digbus.2022.100050
Overton, M., & Kleinschmit, S. (2022). Transforming research methods education through data science literacy. *Teaching Public Administration, 41*(2), 149–169. https://doi.org/10.1177/01447394221084488
Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*. https://doi.org/10.1136/bmj.n160. n160.
Pangrazio, L., & Sefton-Green, J. (2020). *The social utility of 'data literacy'. Learning, Media and Technology, 45*(2), 208–220. https://doi.org/10.1080/17439884.2020.1707223
Papamitsiou, Z., Filippakis, M. E., Poulou, M., Sampson, D., Ifenthaler, D., & Giannakos, M. (2021). Towards an educational data literacy framework: Enhancing the profiles of instructional designers and e-tutors of online and blended courses with new competences. *Smart Learning Environments, 8*. https://doi.org/10.1186/s40561-021-00163-w, 18.
Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods, 10*(3), 330–342. https://doi.org/10.1002/jrsm.1354
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*. https://doi.org/10.1186/s40594-024-00484-5, 25.
Raffaghelli, J. E. (2019). Developing a framework for educators' data literacy in the European context: Proposal, implications and debate. In *Proceedings of the 11th international conference on education and new learning technologies (EDULEARN19)* (pp. 10520–10530). Palma, Spain: IATED. https://doi.org/10.21125/edulearn.2019.2655.
Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report [Report]*. SSHRC Knowledge Synthesis. https://doi.org/10.13140/RG.2.1.1922.5044. Retrieved from https://dalspace.library.dal.ca//handle/10222/64578. (Accessed 17 August 2025).
Schield, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly, 28*(2–3), 6. https://doi.org/10.29173/iq790
Schüller, K., Busch, P., & Hindinger, C. (2019). *Future skills: Ein framework für data literacy*. Berlin: Hochschulforum Digitalisierung [Working Paper No. 47]Retrieved from https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_47_DALI_Kompetenzrahmen_WEB.pdf. (Accessed 17 August 2025).
Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2*(1), 41–63. https://doi.org/10.1007/BF03546488
Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (2005). Windows into the mind. *Higher Education, 49*(4), 413–430. https://doi.org/10.1007/s10734-004-9448-9
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x
Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics, 12*(3), 9–26. https://doi.org/10.15353/joci.v12i3.3275
Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*(1), 45–55. https://doi.org/10.1057/ejis.2011.51