Computers in Human Behavior Reports 17 (2025) 100583

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Filling the gap in K-12 data literacy competence assessment: Design and initial validation of a questionnaire

Belén Donate-Beby [a,*], Francisco José García-Peñalvo [a], Daniel Amo-Filva [b], Sofía Aguayo-Mauri [b]

[a] Informatic Department, University of Salamanca, Salamanca, 37008, Spain
[b] Informatic Department, University Ramón Llul, La Salle, 08022, Barcelona, Spain

ARTICLE INFO

Keywords:
Data literacy
Learning analytics
Education
Questionnaire
Teacher professional development
21st century abilities

ABSTRACT

As the integration of AI-powered technologies in education grows, data literacy has become a key competence for educators, shaping their ability to navigate and utilize vast amounts of educational data. This study details the development of the Educators Data Literacy Self-Assessment (EDLSA), a questionnaire designed to assess perceived data literacy among K-12 teachers, focusing on its behavioural implications. The development of the EDLSA was rigorous. It involved an exhaustive qualitative review of frameworks and a pilot test in a teachers' Spanish sample ($n = 66$) provided relevant insights for refining the instrument. Finally, we conducted a comprehensive statistical analysis, which confirmed the instrument's robust reliability ($\alpha = 0.976$) in measuring teachers' data management competence. The results of the factorial analysis in piloting primary and secondary education samples led to the readjustment of the proposed dimensions into three categories: comprehensive educational analytics, educational problem-solving through data, and promoting meta-learning students through data and ethical implications. Stemmed from the assessed competencies, the EDLSA instrument provides a comprehensive understanding of the human-computer interaction over data in educational settings. Overall, this self-assessment tool presents robust psychometric properties and a framework definition that paves the way for further development among teachers and researchers.

1. Introduction

1.1. Background

The restrictions imposed by the COVID-19 pandemic in 2020 provoked a paradigm shift in data utilization, forcing education to adapt to the evolving technological landscape (Drouin et al., 2020; Lisitsa et al., 2020). This unprecedented event accelerated the digitalization of educational institutions and data collection processes, leading to the widespread adoption of digital pedagogies and educational analytics (García-Peñalvo & Corell, 2020). As a result, high schools and schools have encountered significant challenges in aligning their practices with technological advancements (Elkordy & Iovinelli, 2021).

The broader societal trend of datification, defined as the increasing use of data and algorithmic manipulation across all domains, has reshaped personal and professional spheres (Malik, 2020; Raffaghelli, 2022). As well as an opportunity for innovation, datification in education has introduced challenges such as bias in AI-driven systems and concerns over data privacy (Malik, 2020; O'Neil, 2016).

The rise of Artificial Intelligence (AI) and the widespread use of big data have significantly transformed both societal and educational decision-making (Alier et al., 2024; Lampou, 2023; Sharma et al., 2023, pp. 127–134). These advancements highlight the growing need for competencies such as AI literacy and data literacy to navigate these complex systems effectively.

In the educational context, key milestones illustrate this rapid evolution in data production (Haenlein & Kaplan, 2019), highlighting the integration of machine learning into learning management systems through adaptive learning platforms in the mid-2000s, the establishment of learning analytics as a critical research field and practice in the 2010s (Ferguson, 2012; Siemens & Baker, 2012, pp. 252–254), and the public release of ChatGPT by OpenAI in November 2022 (Brandl & Ellis, 2023), marking a turning point in the public adoption while sparking widespread discourse on the role of AI in education.

* Corresponding author. Paseo de Canalejas, 169, 37008, Salamanca, Spain.
E-mail addresses: belendonate@usal.es (B. Donate-Beby), fgarcia@usal.es (F.J. García-Peñalvo), daniel.amo@salle.url.es (D. Amo-Filva), sofia.aguayo@salle.url.edu (S. Aguayo-Mauri).

https://doi.org/10.1016/j.chbr.2024.100583
Received 16 September 2024; Received in revised form 5 December 2024; Accepted 28 December 2024
Available online 3 January 2025
2451-9588/

These advancements can potentially enhance personalization in learning experiences but also raise significant concerns (Carretero et al., 2017; COMEST, 2019; UNESCO, 2022).

Diverse research areas have increasingly focused on conceptualizing the competencies required for citizens and professionals, mainly on AI literacy, defined by understanding AI systems and recognizing their limitations and ethical implications (Selwyn et al., 2020). Within this context, AI literacy, focused on understanding AI systems and addressing their ethical implications, aligns closely with the Safe AI in Education Manifesto principles (Alier-Forment et al., 2024). The Manifesto extends AI literacy by emphasizing secure, ethical, and transparent AI integration in education, ensuring alignment with pedagogical goals and addressing issues like privacy, bias, and accountability (Alier et al., 2025; Peñalvo et al., 2024).

Amidst these advancements, data literacy has emerged as a critical competency, addressing the need for educators to effectively engage with data and AI-driven systems in educational contexts (Olari & Romeike, 2021). Data Literacy coincides with AI literacy in the understanding that computing systems learn from data, critically interpreting and evaluating AI systems' input and output data (Long & Magerko, 2020, pp. 1–16; Markham et al., 2018). In other words, Data literacy is a crucial competence for AI users, as data serves as an essential resource for AI systems.

However, gaps persist in addressing the intersection of AI literacy and data literacy, particularly in empowering educators to leverage data for pedagogical innovation. Despite the growing use of AI in education, the process remains complex, involving data collection, cleaning, tagging, analysis, and reporting (Amo-Filva et al., 2023). Addressing these complexities requires educators to emphasize ethical considerations, data privacy, and security concerns (Amo-Filva et al., 2021; Amo-Filva et al., 2023; Amo-Filva et al., 2023).

Effective interaction with data has become essential for informed participation in modern society (Salomão Filho et al., 2023; Vuorikari et al., 2022). Data literacy, defined as the ability to collect, manage, evaluate, and critically apply data (Ridsdale et al., 2015), intersects with broader constructs such as digital, information, and statistical literacies. The evolution of data literacy is rooted in statistical literacy concept, with early discussions focusing on the evaluation of information as a key element (Biezā, 2020; Shields, 2005). Over time, the concept has expanded to encompass broader skills needed in a data-driven society (Wolff et al., 2016). Thus, unlike information literacy, which focuses on the broad spectrum of information management, data literacy narrows its approach to extracting, interpreting, and visualizing data for specific decision-making purposes (Law et al., 2018). Moreover, data literacy requires a certain level of statistical literacy to perform fundamental tasks such as interpreting graphs and analyzing trends (Ridsdale et al., 2015).

Grounded in digital literacy, data literacy also ensures the safe and efficient use of digital technologies to manage and create data (Buckingham, 2010; Julien, 2018). However, digital literacy emphasizes the technical ability to access and manage information within digital environments (Burke & Mann, 2024; Vuorikari et al., 2022) while data literacy builds on this foundation by focusing on the critical interpretation, evaluation, and communication of data (Wang, 2020).

Data Literacy in K-12 education involves teachers managing and utilizing data in their daily teaching and learning practices. Building on the DigComp 2.2 Framework (Vuorikari et al., 2022) and Ridsdale et al. (2015), this study focuses on competencies related to creating actionable information from data, processing, and evaluating it. These competencies include, from a procedural approach.

- **Evaluating Data:** Critically assess data sources for credibility and accuracy, understand biases in data collection, and apply ethical principles in using student-related data (Ridsdale et al., 2015).
- **Organizing and Processing Data:** Structure, store, and retrieve data to support decision-making processes in education (Maybee & Zilinski, 2015).
- **Interpreting and Visualizing Data:** Using tools to transform raw data into actionable insights, such as creating graphs and analyzing trends to improve instructional strategies (Mandinach et al., 2016).

Furthermore, the study explores specific categories of data commonly used in K-12 education, aligning them with tasks and skills identified in the literature (Ridsdale et al., 2015; Amo-Filva, Campión, & Prieto, 2017; Wolff et al., 2019).

- **Student Behavioral Data:** This category includes information on attendance, participation, homework completion, and other observable behaviors during learning activities. Such data provides insights into student engagement and informs adjustments to teaching strategies.
- **Performance and Assessment Data:** This type of data encompasses results from both formal assessments (e.g., tests, quizzes, standardized exams) and informal evaluations (e.g., classroom observations, activities). It is instrumental in evaluating student outcomes and the efficacy of instructional approaches.
- **Learning Analytics Data:** Derived from digital learning platforms like Learning Management Systems (LMS), this data includes clickstream patterns, submission logs, and time-on-task metrics. These analytics enable educators to identify trends in student learning behaviors and customize their teaching methods (Siemens, 2013).
- **Meta-Learning and Ethical Data:** This dimension highlights educators' reflection on the ethical use of data (Flores-Vivar & García-Peñalvo, 2023), emphasizing privacy and security when managing sensitive information, such as student demographics or personal details.

The growing use of AI and data produced in educational institutions (Abdullah, 2022) does not align with the existing AI literacy frameworks, which overlook data literacy in understanding AI (Olari & Romeike, 2021). This gap highlights the importance of explicitly integrating data literacy competencies into educational frameworks, particularly in response to the increasing prevalence of AI-driven practices in schools. In line with this, UNESCO's (2022) has recently developed knowledge outcomes and skills mapping for elementary, lower, and upper secondary education, supported by 11 governments implementing 14 AI curricula with data literacy activities.

Regarding elementary education, this report has identified an understanding of data trends in students' performance, the principles and processes of data collection, and straightforward analysis. In lower and upper secondary education, the findings emphasize a curriculum that includes understanding data collection methods, data processing, analysis, reporting, identifying data sources, and the ability to describe the basic structure of a table in a spreadsheet.

In upper secondary education, more advanced tasks are identified, such as describing data and information characteristics, evaluating storage management capabilities, understanding structured and unstructured data, exploring coding techniques for efficient data representation, and visualizing large datasets using visualization tools. When considering data literacy within the educational context, especially in relation to evaluation and critical thinking skills, it is essential to understand it as the ability to analyze and interpret various data types by integrating quantitative data with curricular, pedagogical, and disciplinary knowledge (Mandinach & Gummer, 2015). This cyclical process of monitoring, evaluating, and applying improvements through critical thinking (Bowers, 2017) supports the teaching-learning process by analyzing procedural data and its effects on student performance (Mandinach & Abrams, 2022; Zaytseva et al., 2019). Moreover, data literacy is not only a technical competence but also significantly impacts daily teaching practice (Henderson & Corry, 2021; Wolff et al., 2016).

Concretely, in the K −12 educational context, providing teachers with the necessary support in implementing data-driven methods is crucial because they can be complex and unfamiliar to them (Hegestedt et al., 2023; Melnikova et al., 2023), being essential to balance teachers' technical capacity and adequately sophisticated data products (Gummer & Mandinach, 2015).

Our research aims to address this gap by providing a tool for educators and researchers through a initially validated quantitative questionnaire that assesses data literacy adjusted to K-12 teachers' behaviour and competencies. This in process validated questionnaire could help teachers understand and be aware of their current level of data literacy (J. Raffaghelli, 2019) and guide their professional development in this area (Caena & Redecker, 2019b).

1.2. Problem statement and objectives

Current frameworks and concepts related to data literacy do not provide a cohesive interpretative framework for systematic application in real educational settings (Wise & Shaffer, 2015).

Thus, the problem statement is described by the absence of a consensus framework for the particular characteristics and needs of K-12 teachers, together with the non-existence of cross-validated instruments for measuring data literacy in education (Cui et al., 2023; Ghodoosi et al., 2023). It highlights the need to analyze the theoretical evidence developed so far and its dimensional feasibility in teaching practice at these stages. The following criteria serve as a reference for formulating and developing the questionnaire: (1) choosing the most suitable framework for learners and every educational context that closely matches teachers' characteristics and needs, typically represented in these educational stages; (2) testing the dimensional characteristics presented at the theoretical level through a pilot study on a sample of K-12 teachers.

This study addresses gaps in existing frameworks by developing, piloting, and initially validating the Educators' Data Literacy Self-Assessment (EDLSA) questionnaire. This tool evaluates K-12 teachers' self-perceived data literacy competencies, providing a foundation for conducting research and practical applications of the questionnaire in future studies. Thus, in line with these objectives, the research seeks to address the following questions.

- **RQ1**: How do existing data literacy frameworks address K-12 educators' specific needs concerning data-driven primary and secondary education curricula?
- **RQ2:** How can a self-assessment tool be designed to measure K-12 teachers' self-perceived data literacy competencies accurately?
- **RQ3:** What are the results of the initial pilot testing of the EDLSA questionnaire, and how do these findings inform the confirmatory factor analysis (CFA) for future validation?"

2. Methodology

Following the research objectives, we defined the research design as a sequential exploratory design (DEXPLOS) with a derivative methodology (Creswell, 2012).

For the statistical analysis from the quantitative questionnaire validation, we have followed the recommendations by López et al. (2015) and the guide proposed by Twining et al. (2017) for developing the qualitative methods. As Fig. 1 displays, this section describes the steps taken for the questionnaire development process.

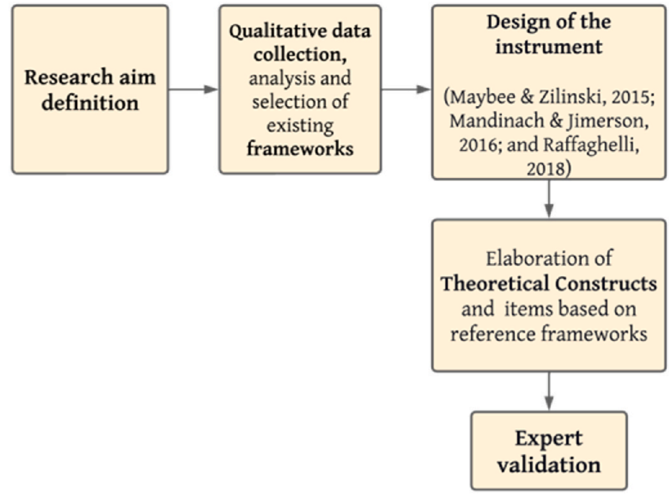

Fig. 1. Work status diagram.

Thus, the method consists of three sub-sections.

- One section contextualizing the work progress in the questionnaire theoretical development, including the instrument design and the expert validation.
- The dimensions proposed theoretically need to be tested in a primary and secondary education sample. Thus, the second section presents the steps taken for the final questionnaire development, including the pilot study, the sampling and statistical validation methods.
- The third comprises the validation process, demonstrating statistical analysis methodology.

2.1. Framework selection

To conceptualize the questionnaire, we first conducted a documentary analysis using a qualitative methodology. This process involved reviewing relevant frameworks and transforming them into specific activities, which served as the foundation for item development. Following item creation, an expert panel evaluated the reliability and relevance of each item (see 3.1.2 Expert Validation; Donate et al., 2022). Thus, this section provides a detailed overview of the methodology used for the documentary analysis (Bowen, 2009). The databases for the search process were WoS and Scopus, as they met the articles' quality requirements.

Given the urgent need for a unified framework in K-12 data literacy

Table 1
Criteria for selecting data literacy frameworks for K-12 education.

Criteria	Inclusion	Exclusion
Focus Area	Data literacy frameworks for general education.	Digital literacy frameworks without a focus on data in education.
Educational Level	Frameworks intended for K-12 or general educational settings.	Frameworks specific to higher education or corporate environments.
Publication Date	Published within the last 10 years.	Published over 10 years ago.
Quality	Peer-reviewed sources from high-impact journals.	Non-peer-reviewed or non-academic sources.
Alignment with Key Models	Frameworks aligned with recognized data literacy models (e.g., Mandinach & Jimerson, Ridsdale et al.).	Frameworks focused on technical/statistical data skills only.

(Henderson & Corry, 2021), we conducted a documentary analysis to identify and evaluate relevant frameworks (Laupichler et al., 2023). This process was informed by specific criteria (Table 1).

The criterion on "Focus" ensures the inclusion of frameworks that specifically address data literacy, excluding those focused on broader digital skills that might dilute the analysis. This targeted Focus aligns with the study's goal of identifying data-specific competencies required in K-12 education. Frameworks were further selected based on their applicability to K-12 education, recognizing the unique developmental needs of teaching this age group. Excluding frameworks tailored explicitly for higher education or corporate settings ensures relevance to the technical levels and requirements of K-12 educators and students.

Only frameworks published within the past ten years were included to capture current trends and standards. Additionally, only peer-reviewed frameworks from reputable sources were considered to ensure rigour. Aligning with established data literacy models (e.g., Mandinach & Jimerson, Ridsdale et al.) was essential for enhancing consistency across selected frameworks and identifying core data literacy competencies relevant to K-12 education while excluding overly technical or specialized competency statistical models.

The final selection of frameworks includes those by Buckingham Shum (2012), Ridsdale et al. (2015), Maybee and Zilinski (2015), Mandinach and Jimerson (2016), Redecker and Punie (2017), and Raffaghelli (2018).

Finally, we have developed the rationale that integrates the final framework selection (View Fig. 2.).The principal reference framework selected was one proposed by Mandinach and Jimerson (2016) and Raffaghelli (2018), because they tailored their theoretical proposal to consider the challenges presented in the educational landscape. These authors founded their proposal by embracing the basis of later research, highlighting Maybee and Zilinski (2015). The aim of developing a data literacy framework for higher education in the context of disciplinary learning led these authors to define the elements of generic data literacy models. This proposal, together with the one suggested by Ridsdale et al. (2015), has displayed a wide enough length to establish the theoretical basis of another educational sample, such as K-12 teachers. According to the latter, competencies are split into levels (core, conceptual and advanced), process dimensions in educational data usage based on a large number of authors' theoretical proposals (conceptual framework, data collection, data management, data evaluation and data application) and, in turn, several critical skills displayed in Fig. 2. As can be seen on it, we followed these dimensions that better fit in complexity and in the tasks that generally would be developed in these levels.

As the axis that operates the actions and competencies proposed in

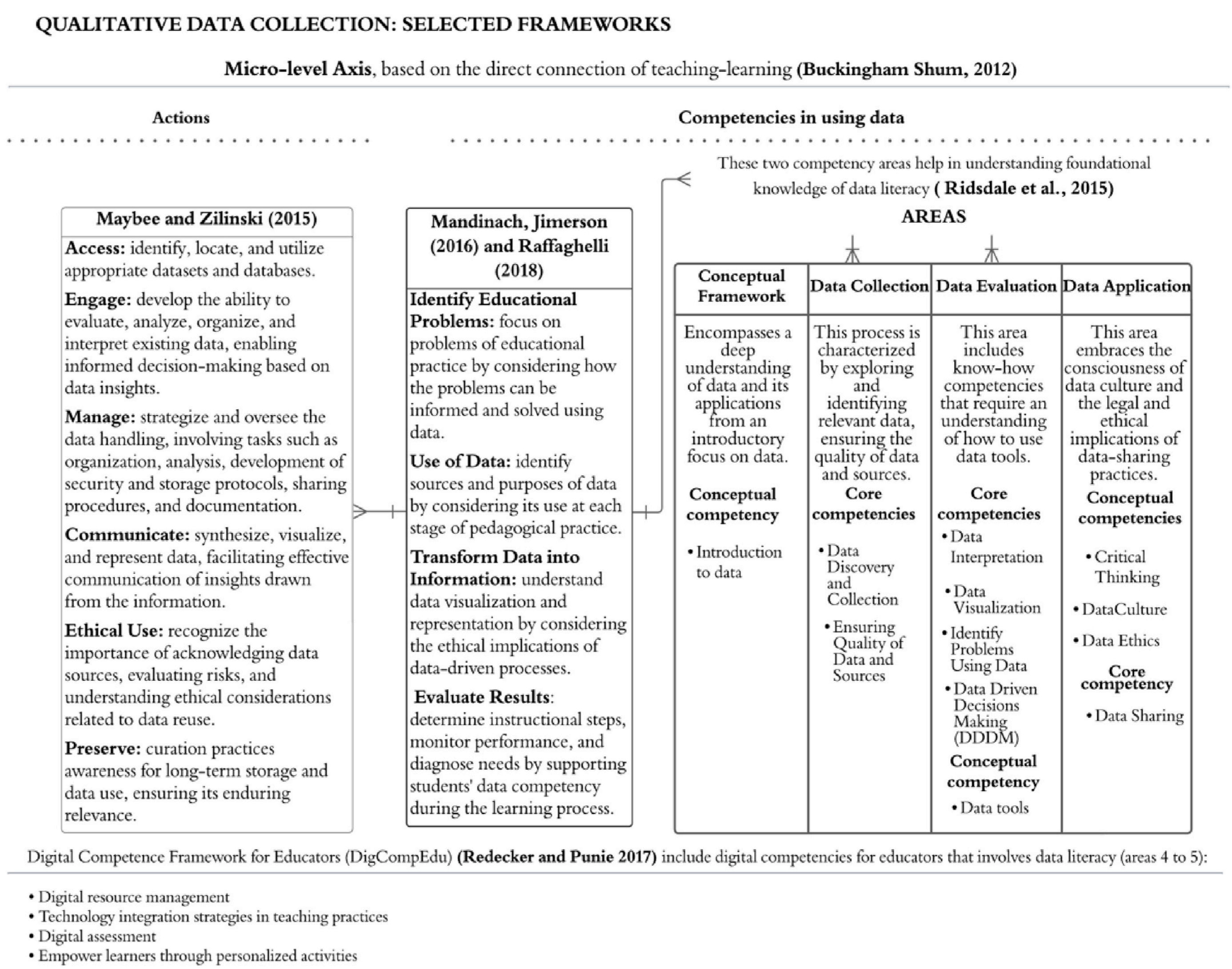

Fig. 2. Selected frameworks: Buckingham Shum's (2012), Ridsdale et al. (2015), Maybee and Zilinski (2015), Mandinach and Jimerson (2016), Redecker and Punie (2017) and Raffaghelli (2018).

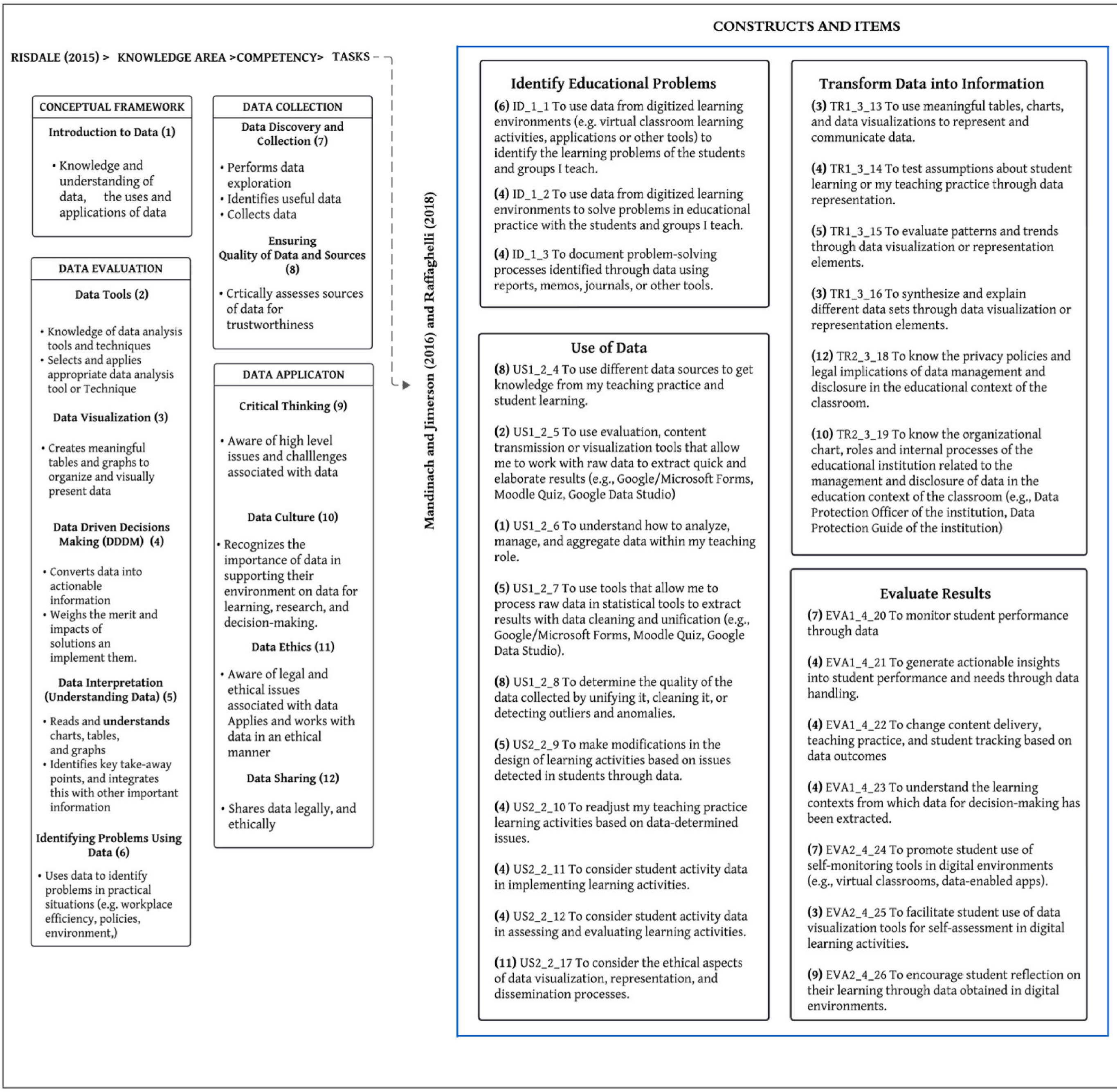

Fig. 3. Constructs and items development through the theoretical frameworks Mandinach and Jimerson (2016), Raffaghelli (2018) and Ridsdale et al. (2015) proposal.

the frameworks, we adopted the work of Buckingham Shum's (2012) work to specify the educational action type and contextualize the questionnaire items. This coincided with the Micro-level, based on the direct connection of teaching, and the Meso-level, which operates through policies and actions that influence the Micro. Furthermore, we used the proposal of Redecker and Punie (2017) to generate items based on the current European Union Competence Framework for Educators.

- Once the frameworks were meticulously established, we developed the constructs by creating theoretical relations between them as shown at Fig. 3:
- The tasks and knowledge proposed by Ridsdale et al. (2015), were taken to develop the items from each dimension.
- Concerning Ridsdale et al. (2015) proposal, to suit the items to the sample characteristics, we remove advanced competencies and those competency dimensions that involve the development of tasks of a higher complexity than expected in the primary and secondary school context, discarding the 'data management' dimension.

2.1.1. Expert validation

Afterward and as the last step, a group of experts in data literacy overlap in various suggestions to improve the validity and reliability of the questionnaire. The group was comprised by twenty teachers and ten post-graduate students. As their expert conclusions in the adequacy of the constructs measured, we proceed to.

- Compose the phrasing of the items in the infinitive form, adding examples and simplifying or elaborating on technical concepts that required clarification.
- Develop a Likert-type scale and split into two sections to measure perceived ability (1–4) and current practice (5–6). In this line, findings indicate that having more scale points, as the 6 points-scale, seems to reduce skewness, and promote the normal sample distribution (Leung, 2011).

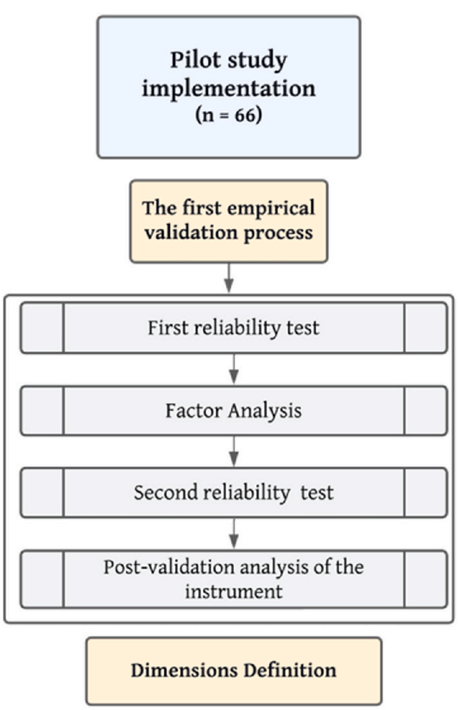

Fig. 4. Study design and data analysis workflow.

2.2. First empirical validation

The process of validating a questionnaire aims to demonstrate its feasibility, reliability, accuracy, and suitability for the problem it intends to measure (content validity), being sensitive to changes in responses and reflecting the underlying theory of the phenomenon or concept it aims to measure (construct validity) (de Yébenes Prous et al., 2009).

The software used to carry out all the statistical processes was IBM SPSS 28 (SPSS, INC., Chicago, IL) (George & Mallery, 2019), and the stages for the validation process are presented in Fig. 4 and are outlined below.

2.2.1. Pilot study

This pilot study aims to determine whether the research design, methods, and data collection tools are practical and can be implemented effectively on a larger scale (Hazzi & Maldaon, 2015). For this reason, we obtained initial data to provide insights into the potential outcomes of a future full-scale study. Consequently, this pilot study is expected to consist of a first tool validation, ensuring the effectiveness of the research approach.

This pilot study was conducted online from March to May 2023. The questionnaire was administered online through Google Forms, and to encompass possible privacy issues, the email used for the dissemination process was institutional from the document anonymization. Thus, its dissemination was developed through a teacher of each institution.

2.2.2. Sample

In this first empirical instrument validation, the questionnaire was conducted to a sample of 66 teachers, 24 primary school teachers, and 48 secondary school teachers, using the snowball sampling technique as the selection methodology. Snowball sampling is a non-probability method used to identify and recruit participants with homogeneous features, facilitating access to limited-access populations (Naderifar et al., 2017).

For this first empirical approach, personal data, such as the genre or age, were not collected. Also, participants belong to various educational Spanish institutions from Canary islands, Madrid, Asturias, Cataluña, and Oviedo. Yet, for privacy preservation, this information has not been collected individually. The exclusion and inclusion sample criteria are shown in Table 2.

2.3. Statistical analysis

2.3.1. Data curation and exploration

To facilitate exhaustive statistical analysis, the dataset was preprocessed and cleaned. We excluded the information that did not contribute to the initial statistical validation. Therefore, the dataset exclusively comprises the Likert-type scale results.

We initiated the data exploration by analyzing the responses' central value and dispersion. This preliminary analysis also involved examining the relationships between the items and their strengths through the matrix of correlations of the variables (Oviedo & Campo-Arias, 2005). This matrix highlighted high indices, indicating strong relationships between the items.

Table 2
Inclusion and exclusion criteria.

INCLUSION CRITERIA	EXCLUSION CRITERIA
Currently teaching.	Not currently teaching.
Teaching in Spanish schools and high schools.	Teaching in countries distinct from Spain.
Teaching in Elementary and Secondary Education institutions.	Teaching in other institutions.
Teaching in private or public formal education institutions.	Teaching in other non-formal education institution.

2.3.2. First reliability test

Cronbach's alpha coefficient allowed us to determine the questionnaire's reliability and internal consistency, ranging between 0 and 1 (Kılıç, 1970). This test returned high indices in the consistency of individual items. To continue, we conducted a second analysis to verify the results by dividing the sample into two parts using the two-half technique (Split-Half Analysis). Concretely, the test conducted was the Spearman-Brown corrected reliability test (Feldt & Charter, 2003). In their study, Wadkar et al. (2016), established a scale based on Cronbach's Alpha coefficient to assess the instrument's reliability. According to the interpretive guidelines for reliability coefficients, a coefficient of *0.90 or higher* is considered *excellent, 0.81 to 0.90* is *good, 0.71 to 0.80* is *acceptable, 0.61 to 0.70* is *questionable, 0.51 to 0.60* is *poor*, and *below 0.50* is *unacceptable*. The test results revealed a reliability coefficient of 0.96, which falls within the *excellent* range, indicating a high level of reliability for the measure. Subsequently, we performed an item-deleted alpha test to preliminary analyze the weight of each question for the questionnaire's reliability. The if-item-deleted alpha analysis showed how alpha would change if a question were not on the test; a lower result means that the question should not be considered deletion because it would lower the overall alpha (Streiner, 2003).

2.3.3. Exploratory Factor Analysis (AFE)

We executed an Exploratory Factor Analysis (AFE) to assess the questionnaire's theoretical constructs (Glass & Arnkoff, 1997). The EFA aimed to identify variables that exhibit shared meanings, allowing us to comprehend and analyze the structural interrelationships among variables pertinent to managing educational data in schools and high schools.

Before the analysis, we used the Kaiser-Meyer-Olikin index (KMO) to verify dataset adequacy for the EFA. The KMO index allows for assessing the magnitude of partial correlations among the variables. In that line, Kaiser's Rule (Braeken & Van Assen, 2017) suggests that only these factors with eigenvalues greater than one should be retained. Thus, the results exhibited adequate internal consistency for conducting EFA with a KMO of 0.905.

The extraction method utilized was a sequential approach based on Principal Component Analysis (PCA) development and the Exploratory Factor Analysis (AFE) with the Maximum Likelihood method, aimed at producing a linear combination that explains the highest possible percentage of variance to account for at least 70% of the total variance. Thus, the total variance explained per component matrix was calculated to check whether the data had suitable features for this analysis (Frías-Navarro & Pascual Soler, 2012). Together, the three components explained 78.95% of the total variance in the data. Therefore, the gathered data have displayed suitable features for the EFA.

Nevertheless, as these criteria tend to overestimate the number of factors, we also inspected the scree plot (Cattell, 1966) identifying the inflexion point where the slope of the line connecting the ordered factors no longer decreases. We included only the factors preceding this point in the final analysis.

A Factor Rotation was performed following the Varimax rotation method with Kaiser normalization to improve the interpretation of the factor structure. In this line, the Rotated Component Matrix Variance Component highlighted the pattern of correlations and the factor weights, facilitating the interpretation of the results. As the saving variables method, the Barlett estimation technique based on analysis of variance (ANOVA) was chosen. Using this variable-saving method with this size sample, we ensure that factorial scores are unbiased and that the estimated scores align, on average, with the actual factorial scores (Saris et al., 1978).

We established criteria for interpreting variable saturation, a minimum value of 0.6 for inclusion in a factor. Providing a variable was present in two distinct factors; it will be incorporated into the factor where it holds the highest weight. Subsequently, upon determining the variables allocated to each factor, they were named accordingly.

2.3.4. Second reliability test

The validation process concluded with a second Cronbach Reliability Analysis test to analyze the reestablished dimensions' weights and normal distribution tests. Regarding the latter, a small percentage of data distributions exhibit expected Skewness and Kurtosis values under normality (Blanca et al., 2013). Thus, we have considered a moderate flattening level in the sample acceptable.

2.3.5. Post-validation analysis of the instrument

In this research, the distribution of the relevant statistic is unknown because the population's mean and variance are unidentified, being necessary to check the sample normality. The Kolmogorov-Smirnov and Shapiro-Wilk normality tests were used to assess data normality (Rosenthal, 1968).

Through the statistical hypothesis testing, we have compared the significance probability value (p-value) returned from these tests to determine whether or not to reject the null hypothesis and conclude if the result is statistically significant (Kwak, 2023). Thus, we hypothesized that the result would be statistically significant when $p < 0.05$ and the calculated significance probability is 0.005.

2.3.6. Dimensions definition

Finally, the dimensions were defined, with the items redistributed along the constructs. Regarding the considerations in the research aim definition, we took the theory of response to the item from Bandura et al. (1999), which outlined that a person's internal self-beliefs are predictors of their capabilities. In that line, assessing methods of endorsing self-statements are widely used, and examining internal consistency represents the most suitable approach for evaluating reliability (Glass & Arnkoff, 1997).

3. Results

3.1. Preliminary analysis

A preliminary exploration of response data was conducted by analyzing central tendency and dispersion. Mean scores for each item ranged between 3.00 and 4.00 on a 6-point scale ($M = 3.50$, $SD = 0.75$), indicating a medium to high perceived ability level. The analysis of item response dispersion was assessed using the standard deviation of means, revealing moderate variability across items. Descriptive statistics and a correlation matrix were calculated to evaluate relationships among variables (Oviedo & Campo-Arias, 2005; See Appendix Fig. 5). Correlations were classified as very high (≥0.9), high (0.7–0.899), medium (0.4–0.699), and medium-low (0.1–0.399), with the majority of items displaying moderate to high associations.

3.2. First reliability analysis: Cronbach's alpha results

The Cronbach's Alpha test, complemented by a split-half analysis, revealed high-reliability coefficients, with an overall score of $\alpha = 0.977$ and scores above $\alpha = 0.95$ in both halves. These values signify excellent internal consistency for the questionnaire (see Table 3).

The high values obtained in the consistency test indicate that the variables contribute to the overall measurement of each construct. Furthermore, in the "Cronbach's Alpha if Item Removed" test, we found that although the indices still show strong values, removing items ID_1_1, ID_1_2, and ID_1_3 would lead to a lower Cronbach's Alpha.

Table 3
Reliability statistics.

CRONBACH'S ALPHA	
ALL VARIABLES 26 variables	0.977
PART 1 13 variables	0.957
PART 2 13 variables	0.970

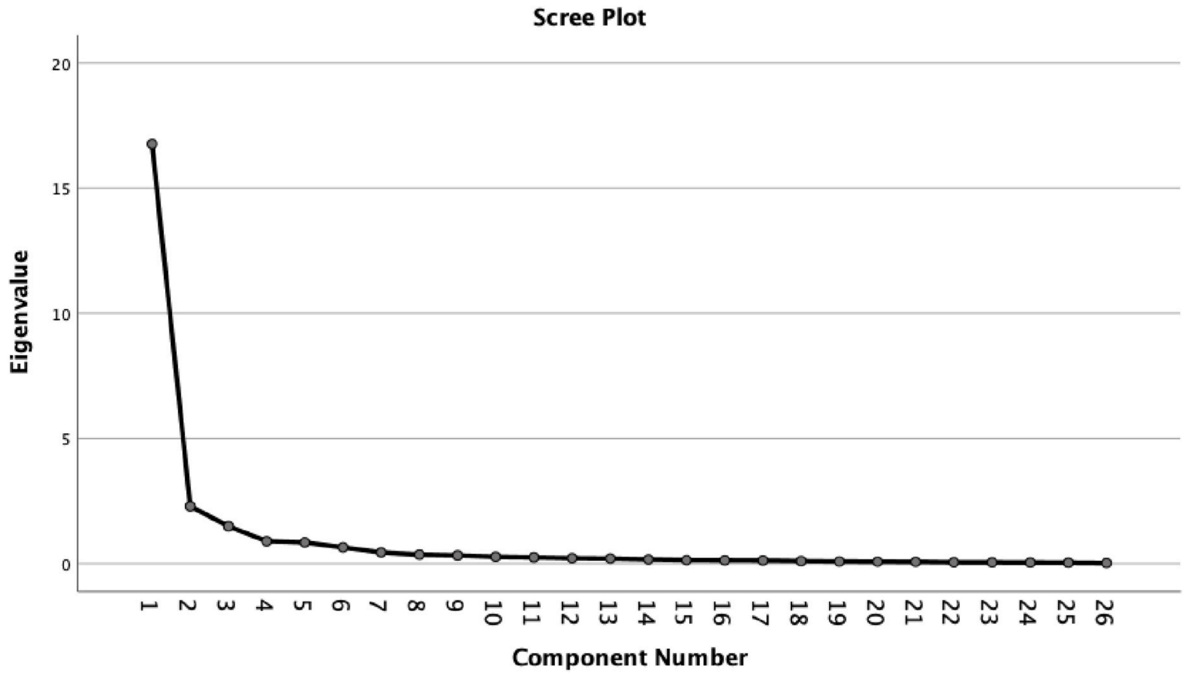

Fig. 6. Scree plot graph.

Table 4
Total variance explained aplyied maximum likelihood extraction method.

Factor	Initial eigenvalues			Sums of squared extraction charges			Sums of loads squared by rotation		
	Total	% of variance	% accumulated	Total	% of variance	% accumulated	Total	% of variance	% accumulated
1	16,230	**62.423**	62.423	15.815	60.826	60.826	7.059	27.149	27.149
2	2.255	**8.672**	71.094	1.443	5.552	66.378	6.763	26.012	53.161
3	1.543	**5.936**	77.030	2.035	7.826	74.204	5.471	21.043	74.204
4	1.000	3.848	80.878						
5	0.882	3.394	84.272						
6	0.733	2.821	87.092						
7	0.577	2.220	89.312						
8	0.359	1.379	90.691						
9	0.355	1.365	92.056						
10	0.314	1.206	93.262						
11	0.262	1.006	94.268						
12	0.233	0.897	95.165						
13	0.207	0.795	95.960						
14	0.178	0.686	96.646						
15	0.132	0.506	97.152						
16	0.129	0.497	97.649						
17	0.121	0.466	98.115						
18	0.098	0.379	98.494						
19	0.081	0.312	98.806						
20	0.075	0.290	99.096						
21	0.067	0.259	99.355						
22	0.047	0.182	99.537						
23	0.038	0.147	99.684						
24	,034	0.130	99.814						
25	,031	0.119	99.934						
26	,017	0.066	100.00						

Hence, it is required to retain these items.

To identify whether this questionnaire has sufficient internal consistency to carry out the factor analysis, the Correlation between the two halves and the Spearman-Brown corrected reliability test has been carried out. The test returned $\alpha = 0.907$, which underlines the high reliability of the scale.

3.3. Exploratory Factor Analysis (EFA)

3.3.1. Factor Analysis suitability

The significance of factor weights was assessed using Kaiser-Meyer Olkin (KMO) sampling adequacy index. The KMO value obtained was 0.905, indicating that the sample is remarkably adequate. In this case this report suggest that the factor analysis might be appropriate.

Table 5
Rotated factor matrix with maximun likelihood extraction method.

Rotated Factor Matrix			
	Factor		
	1	2	3
ID_1_1	0.204	**0.655**	0.183
ID_1_2	0.203	**0.772**	0.190
ID_1_3	0.082	**0.804**	0.337
US1_2_4	0.286	**0.676**	0.304
US1_2_5	0.365	**0.726**	0.191
US1_2_6	0.370	**0.755**	0.254
US1_2_7	0.434	0.595	0.317
US1_2_8	0.504	0.577	0.242
US2_2_9	0.318	**0.749**	0.328
US2_2_10	0.314	**0.772**	0.216
US2_2_11	**0.694**	0.430	0.419
US2_2_12	**0.682**	0.434	0.427
TR1_3_13	**0.749**	0.207	0.357
TR1_3_14	**0.783**	0.268	0.366
TR1_3_15	**0.790**	0.235	0.391
TR1_3_16	**0.802**	0.276	0.356
TR2_3_17	0.469	0.224	**0.776**
TR2_3_18	0.462	0.297	**0.654**
TR2_3_19	0.412	0.267	**0.639**
EVA1_4_20	**0.673**	0.372	0.412
EVA1_4_21	**0.812**	0.344	0.303
EVA1_4_22	**0.779**	0.385	0.323
EVA1_4_23	−0.264	−0.082	0.014
EVA1_4_24	0.260	0.401	**0.828**
EVA1_4_25	0.244	0.351	**0.859**
EVA1_4_26	0.257	0.410	**0.810**

Extraction method: Maximum Likelihood.
Rotation method: Varimax with Kaiser normalization.
a. The rotation has converged in 7 iterations.

Therefore, the data are suitable for a factor analysis.

3.3.2. Exploratory Factor Analysis (EFA) development

We have employed a sequential approach based on aligning the PCA and EFA results while emphasizing the interpretability of the three-factor solution. The first method used was Principal Component Analysis (PCA), followed by Exploratory Factor Analysis (EFA) using the Maximum Likelihood Extraction Method with VARIMAX Rotation to ensure independence among factors. PCA was applied as a data compression method to simplify the pattern of loadings, improve interpretability, and determine the structure for subsequent analysis.

In the PCA, three components together explained 78.9% of the total variance, leaving 21.1% unexplained. Additionally, this first three components have an eigenvalue >1.00, which accounted for a significant portion of the variance. The PCA rotated matrix weights and the scree plot (View Fig. 6) further confirmed the appropriateness of retaining three components, as the curve levelled off after the third component.

Subsequently, Table 4 details the percentage of variance explained by the three factors in the EFA development with Likelihood Maximum Method. This table presents the initial eigenvalues, sums of squared extraction charges, and sums of squared loadings after rotation. The three-factor solution derived through EFA accounted for 77.03% of the total variance in the data, demonstrating that the structure identified is robust and interpretable.

The Rotated Factor Matrix (Table 5) clarified the correlation patterns among variables, aiding in interpretability. Using a significance threshold of 0.60, discrepancies were observed between the PCA-based component matrix and the Maximum Likelihood-based rotated matrix. While the item US2_2_12 was initially flagged as non-significant in the PCA analysis, its reevaluation during the EFA suggested retention due to acceptable loadings. Conversely, US1_2_7, US1_2_8, and especially EVA1_4_23 failed to load significantly on any factor (Table 5).

Table 6
Second reliability test.

RELIABILITY STATISTICS		
FACTOR	CRONBACH'S ALPHA	NUMBER OF ITEMS
1	0.976	9
2	0.945	8
3	0.963	6
TOTAL	**0.961**	**23**

3.3.3. Dimensions configuration

The Rotated Component Matrix was prioritized for final decisions, as it enhances interpretability and aligns with theoretical constructs. Based on these results, during the Exploratory Factor Analysis (EFA) process, factors were interpreted using a significance threshold of 0.60 for loadings. The analysis revealed that some items clustered into constructs that reflected the underlying dimensions. However, based on the results from the Rotated Component Matrix (Table 5), items US1_2_7 and US1_2_8, while marginally below the 0.60 threshold, were excluded due to their technical complexity in the context of K-12 data literacy. Additionally, item EVA1_4_23 was removed as it does not saturate significantly in any factor.

The decision to remove these items was carefully examined under multiple criteria, including the distribution shown in Fig. 2, the theoretical framework of the items, and the factor weights. The reliability of the constructs was assessed both with and without these items, showing minor differences. Moreover, before finalizing its removal, the theoretical basis for dimensionality (Maybee & Zilinski, 2015; Mandinach & Jimerson, 2016; J. E. Raffaghelli, 2018), the population's adequacy for responding to these items, and their theoretical relevance for measuring the associated dimensions were all considered. Thus, the items removed from the questionnaire were.

- US1_2_7: *To use tools that allow me to process raw data in statistical tools to extract results with data cleaning and unification (*e.g., *Google/Microsoft Forms, Moodle Quiz, Google Data Studio).*
- US1_2_8: *To determine the quality of the data collected by unifying it, cleaning it, or detecting outliers and anomalies.*
- EVA1_4_23: *To understand the learning contexts from which data for decision-making has been extracted.*

3.4. Second reliability analysis

The second Cronbach Reliability Analysis verified that the construct validity and questionnaire reliability are maintained with the new clusters, improving their weights, by analyzing the Conbach's Alpha of each Factor (View Table 6). In addition, we have developed the Average Variance Extracted (AVE) method to evaluate the convergent validity.

3.5. Post-validation analysis of the instrument

A normality test was conducted by examining Kurtosis and Skewness values. Kurtosis values and standard errors indicate moderate flattening of distributions across factors without extreme deviations from normality. Negative kurtosis suggests shorter tails than a normal

Table 7
Kolmorov-Smirnov and Shapiro Wilk normality test.

	KOLMOGOROV-SMIRNOV			SHAPIRO-WILK		
	Statistic	n	Sig.	Statistic	n	Sig.
FACTOR 1	0.072	66	0.200*	0.987	66	0.693
FACTOR 2	0.060	66	0.200*	0.986	66	0.646
FACTOR 3	0.104	66	0.074	0.983	66	0.479

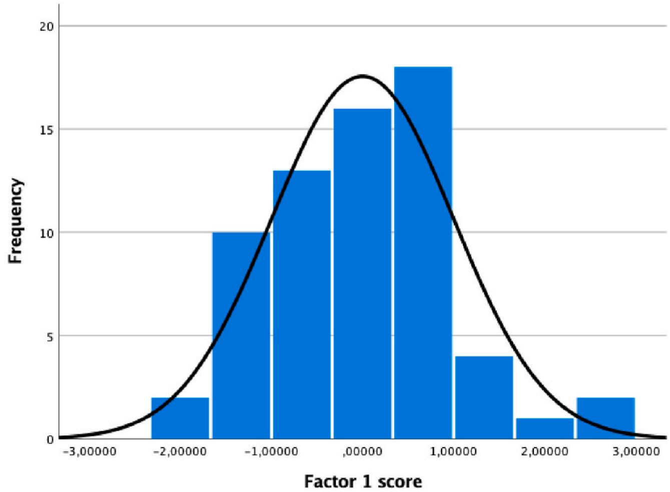

Fig. 7. Factor 1 sample distribution.

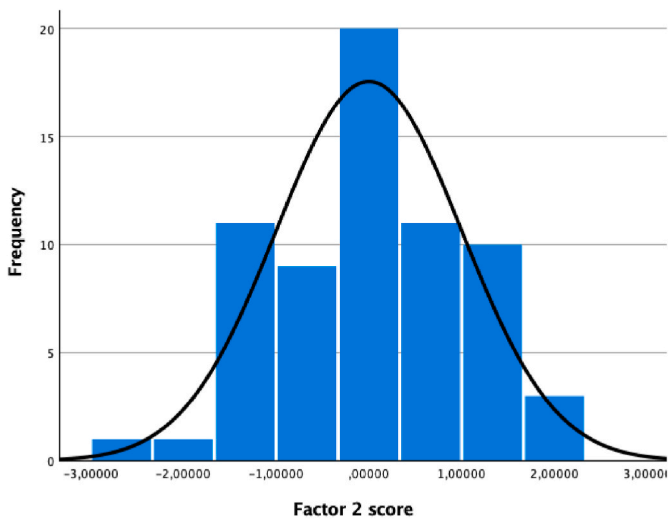

Fig. 8. Factor 2 sample distribution.

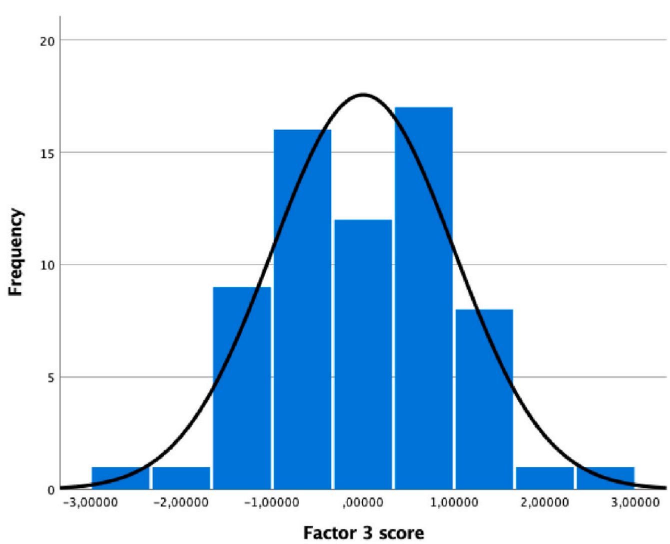

Fig. 9. Factor 3 sample distribution.

COMPREHENSIVE EDUCATIONAL ANALYTICS

- US2_2_11 To consider student activity data in implementing learning activities.
- US2_2_12 To consider student activity data in assessing and evaluating learning activities.
- TR1_3_13 To use meaningful tables, charts, and data visualizations to represent and communicate data.
- TR1_3_14 To test assumptions about student learning or my teaching practice through data representation.
- TR1_3_15 To evaluate patterns and trends through data visualization or representation elements.
- TR1_3_16 To synthesize and explain different data sets through data visualization or representation elements.
- EVA1_4_20 To monitor student performance through data.
- EVA1_4_21 To generate actionable insights into student performance and needs through data handling.
- EVA1_4_22 To change content delivery, teaching practice, and student tracking based on data outcomes.

Fig. 10. Factor 1 comprehensive educational analytics.

EDUCATIONAL PROBLEM-SOLVING THROUGH DATA

- ☐ ID_1_1 To use data from digitized learning environments (e.g., virtual classroom learning activities, applications, or other tools) to identify the learning problems of the students and groups I teach.
- ☐ ID_1_2 To use data from digitized learning environments to solve problems in educational practice with the students and groups I teach.
- ☐ ID_1_3 To document problem-solving processes identified through data using reports, memos, journals, or other tools.
- ☐ US1_2_4 To use different data sources to get knowledge from my teaching practice and student learning.
- ☐ US1_2_5 To use evaluation, content transmission or visualization tools that allow me to work with raw data to extract quick and elaborate results (e.g., Google/Microsoft Forms, Moodle Quiz, Google Data Studio)
- ☐ US1_2_6 To understand how to analyze, manage, and aggregate data within my teaching role.
- ☐ US2_2_9 To make modifications in the design of learning activities based on issues detected in students through data.
- ☐ US2_2_10 To readjust my teaching practice learning activities based on data-determined issues.

Fig. 11. Factor 2 educational problem-solving through data.

distribution, which supports acceptable normality in the sample.

Skewness values are close to 0, indicating approximate symmetry in the distribution. The standard error of skewness was consistent across factors (SE = 0.29), indicating moderate flattening across factors while suggesting a non-extreme departure from normality.

Kolmogorov-Smirnov and Shapiro-Wilk tests (Table 7) confirm normality for all three factors, with $p > 0.05$, leading to the acceptance of the null hypothesis that data are normally distributed.

Additionally, to the outcomes of the normality tests, the ensuing

PROMOTING META-LEARNING STUDENTS THROUGH DATA AND ETHICAL IMPLICATIONS

- ☐ US2_2_17 To consider the ethical aspects of data visualization, representation, and dissemination processes.
- ☐ TR2_3_18 To know the privacy policies and legal implications of data management and disclosure in the educational context of the classroom.
- ☐ TR2_3_19 To know the organizational chart, roles and internal processes of the educational institution related to the management and disclosure of data in the education context of the classroom (e.g., Data Protection Officer of the institution, Data Protection Guide of the institution)
- ☐ EVA2_4_24 To promote student use of self-monitoring tools in digital environments (e.g., virtual classrooms, data-enabled apps).
- ☐ EVA2_4_25 To facilitate student use of data visualization tools for self-assessment in digital learning activities.
- ☐ EVA2_4_26 To encourage student reflection on their learning through data obtained in digital environments.

Fig. 12. Factor 3 Promoting Meta-Learning Students through Data and ethical implications.

histograms visually depict the sample distribution for each dimension. Thus, the distributions of Factor 1 (Fig. 7), Factor 2 (Fig. 8), and Factor 3 (Fig. 9) show relative normality in the distribution of the sample.

3.6. Dimensions definition

The constructs ultimately comprise a total of 23 items. As noted in Figs. 10–12, the factor weights from the matrix of rotated components suggest a new regrouping of the items into constructs. The new dimensional competencies distribution by displaying the relationship between Ridsdale et al. (2015) fundamental competencies and skills/-knowledge areas, Mandinach and Jimerson (2016); Maybee and Zilinski (2015) and Raffaghelli (2018) frameworks and the proposed dimensions for primary and secondary education. Due to the redistribution and the acquisition of a new meaning, new names were assigned to the resulting constructs as follows.

4. Discussion

This study addresses critical challenges and questions regarding data literacy among K −12 educators, aiming to develop a reliable self-assessment tool and explore its effectiveness through initial pilot testing.

> RQ1: How do existing data literacy frameworks address K-12 educators' specific needs concerning data-driven primary and secondary education curricula?

The frameworks such as DigComp 2.1(Carretero et al., 2017) and DigComp 2.2 (Vuorikari et al., 2022) provide key competencies for K-12 educators, particularly in the areas of information literacy and data literacy. While information literacy focuses on the ability to evaluate and use information from various sources, data literacy is more specialized, focusing on data collection, management, analysis, and ethical applications. Educators need to critically assess data accuracy, recognize biases, and ensure that data is used responsibly, especially when dealing with student data and AI-driven educational tools. Information literacy helps in distinguishing reliable sources, while data literacy equips educators with the necessary skills to interpret educational data, fostering a more informed and evidence-based teaching approach (Mandinach & Jimerson, 2016; Ridsdale et al., 2015; Vuorikari et al., 2022).

In addition to DigComp 2.1 and DigComp 2.2, theoretical frameworks developed by Ridsdale et al. (2015), Mandinach and Jimerson (2016), Maybee and Zilinski (2015), and Raffaghelli (2018) further contribute to the understanding of data literacy in education. These frameworks emphasize the importance of critical data skills, focusing on how educators can use data to inform their teaching practices and decision-making. Ridsdale et al. (2015) highlight the importance of data management and evaluation, noting that educators need to understand the types of data they interact with and how to process and make decisions based on that data. Similarly, Mandinach and Jimerson (2016) and Maybee and Zilinski (2015) emphasize that educators must engage with data at various levels—from collecting basic student performance data to using more advanced learning analytics to guide instructional adjustments. Raffaghelli (2018) adds that data literacy should focus on technical competencies and consider the ethical implications of using data in the classroom, ensuring educators are equipped to handle sensitive student information responsibly.

Both competencies are essential for navigating the increasing integration of AI technologies in education, where educators must interact with data and understand how AI systems interpret and use that data. Ethical considerations, particularly regarding privacy and bias, play a central role in data literacy, ensuring that data is used responsibly in the classroom. While DigComp 2.1 and DigComp 2.2 provide a solid foundation for addressing the data-related needs of K-12 educators, further adaptation is needed to fully incorporate the complexities of AI-driven curricula and ensure comprehensive preparation for educators in managing and interpreting data within these new contexts (Maybee & Zilinski, 2015; Vuorikari et al., 2022).

Analytical and evaluative abilities in the digital sphere are crucial for making well-informed decisions and discerning misinformation by assessing source authority, objectivity, and relevance while fostering critical thinking skills to examine information effectively (Naamati-Schneider & Alt, 2024). These competencies are one of the main features of data literacy, which have been theoretically presented for the frameworks developed by the authors Ridsdale et al. (2015), Mandinach and Jimerson (2016); Maybee and Zilinski (2015) and Raffaghelli (2018).

Data literacy has become in an essential teaching competence included in the curricula for AI, data-driven educational practices in the European Union (COMEST, 2019). However, data literacy competencies are still an unexplored avenue for primary and secondary education, and we need tools that allow us to delve deeper into current digital educational practices (Cui et al., 2023).

> RQ2: How can a self-assessment tool be designed to measure K-12 teachers' self-perceived data literacy competencies accurately?

Given the absence of cross-validated instruments for measuring data literacy in education (Ghodoosi et al., 2023), the interest of this work lies in an anonymous self-assessment validated tool aimed at teachers that provides indicators aligned with professional activities. If "teachers can experience their self-assessment process as self-determined and individually owned" (Caena & Redecker, 2019), their predisposition towards developing digital competencies will be before and after.

In response to the need for a reliable measure of data competencies in teaching roles, this study introduces the EDLSA questionnaire, the first validated and anonymous self-assessment tool specifically designed for K-12 educators. To develop this tool, we conducted a comprehensive documentary analysis to define the relevant theoretical constructs for K-12 settings. After an expert group assessment, the tool was piloted with a sample of K-12 educators to ensure its applicability. We provided reliability indicators, analyzed the sample distribution, and conducted Exploratory Factor Analysis (EFA), all demonstrating that the instrument meets sufficient quality standards for broader implementation. These steps highlight its readiness for statistical validation in future research (Laupichler et al., 2023; Schepman & Rodway, 2020).

In today's data-centric world, educators must understand the implications of their digital actions and the data produced through their interactions with students. The EDLSA questionnaire arises from the growing need to understand how data is generated, stored, and utilized in educational environments, particularly as they transition into data-driven settings (Cui et al., 2023).

RQ3: What are the results of the initial pilot testing of the EDLSA questionnaire, and how do these findings inform the confirmatory factor analysis (CFA) for future validation?

According to the design outcomes, the Exploratory Factorial Analysis obtained a dimensional configuration of the items adjusted to the sample. The reliability tests displayed favorable results globally and per factor, with scores close to 1, indicating the questionnaire's construct validity and robustness. Moreover, concerning internal consistency, the new grouping of the items into theoretical constructs seems consistent with their strength and the direction of the relationship between variables (Table 5).

In summary, we observed significantly positive reports on the reliability and internal consistency of the validation process, with a relatively normal distribution of the sample. Even though the results suggest a validated questionnaire designed for full-scale administration, it is necessary to show some of the limitations. The main barriers we found were the complexity of comprehending and reconstructing tasks developed in the teacher's day-to-day work and the control of their implicit meanings of the vocabulary in the questionnaire items development (Amo-Filva, Campión, & Prieto, 2017). In contrast, it is crucial to recognize that the interconnection among theoretically established variables exhibits a heightened complexity in their correlation, as evidenced by the factor analysis conducted in the questionnaire.

Despite the challenges, the questionnaire demonstrated good psychometric properties and a great potential to measure data literacy on the educator's teaching-learning axis from primary and secondary schools. The final version of the questionnaire includes three scales that capture three dimensional characteristics and is available for free use by teachers and researchers (Donate et al., 2024).

However, while relationships between constructs may evolve, the tool's current piloting, demonstrated psychometric properties and comprehensive interpretation of the results. Thus, this work provides evidence of construct validity in measuring key tasks and knowledge surrounding data use, privacy, and decision-making in the classroom. These considerations are crucial as AI and big data technologies increasingly shape the educational landscape, presenting new challenges for educators in managing and interpreting data (Malik, 2020; J. E. Raffaghelli, 2022).

All the dimensions encompass a comprehensive range of questions, some of which have an ad-hoc character, well known as the conceptual competencies presented by Ridsdale et al. (2015), and others present procedural and post-hoc characteristics of the iterative cycle from PPDAC (Wolff et al., 2016). The acronym of this process confers "Plan, Data, Analysis and Conclusion", from which further questions and analyses are generated, often of increasing sophistication as the problem is being solved (White et al., 1999).

The first dimension is *Comprehensive Educational Analytics* (View Fig. 10). This construct refers to collecting and interpreting data, its visualization, information, and evaluating results to optimize the teaching-learning process, requiring pedagogic and contextual knowledge (Amo-Filva et al., 2023) hence, this dimension involves only the micro axis, requiring each teacher's competencies and knowledge for its role activity (Buckingham Shum, 2012). We can also appreciate the items as a shared feature focusing on student learning, formulating pedagogical hypotheses, and corroborating through data. This approach overlaps with the definition of Learning Analytics proposed by Long & Siemens (2011), providing a comprehensive vision of the level of data literacy in education necessary for data-driven decisions. Thus, the main focus of this dimension is data literacy in its educational and pedagogical understanding, which can be considered a sophisticated version of learning analytics in the data-driven decision-making cycle process (Mandinach & Abrams, 2022).

The second dimension corresponds to *Educational Problem-Solving Through Data* (View Fig. 11). Considering problem-solving as a process composed of the evaluation, analysis, identification, and solving of problems through data, this new construct shows items related to the identification, addressing, and communication of educational issues (ID_1_1, ID_1_2, ID_1_, US2_2_9, US2_2_10).

The problem-solving process can be described as 'an individual's capacity to use cognitive processes to resolve real, cross-disciplinary situations where the solution path is not immediately obvious' (PISA, 2003). Likewise, critical thinking and decision-making skills, mentioned before in the *Comprehensive Educational Analytics* dimension, are required for effective problem-solving (Snyder & Snyder, 2008).

In turn, the results indicate that addressing teaching-learning problems may require a higher level of complexity in the use of data through the use of different sources (USI_2_4), the management of raw data (USI_2_5), and the understanding of various data analysis styles, management and aggregation processes (USI_2_6) (Snyder & Snyder, 2008).

The third factor corresponds to Promoting Meta-Learning Students through Data (Redecker & Punie, 2017) and Ethical Implications (Fig. 12), which attempts to measure how teachers involve students in their learning process through data (EVA2_4_24, EVA2_4_25, EVA2_4_26). However, the results indicate that the displayed relationship with students in data management implies ethical, legal, and privacy considerations (US2_2_17, TR2_3_18, TR2_3_19). Therefore, this dimension shows a combination of both competencies: the ability to promote meta-learning in students and knowledge of the ethical, legal, and privacy aspects linked to data in the educational context.

Current frameworks and concepts related to data literacy lack a unified interpretative approach for consistent application in real-world educational settings (Wise & Shaffer, 2015). As a result, the problem arises from the absence of a widely agreed-upon framework that addresses the unique characteristics and needs of K-12 teachers, along with the lack of cross-validated tools to assess data literacy in education (Cui et al., 2023; Ghodoosi et al., 2023).

While relationships between these constructs may evolve, the tool's current piloting, demonstrated a comprehensive interpretation of the results. Thus, this work provides evidence of construct validity in measuring key tasks and knowledge surrounding data use, privacy, and decision-making in the classroom. These considerations are crucial as AI and big data technologies increasingly shape the educational landscape, presenting new challenges for educators in managing and interpreting data (Malik, 2020; J. E. Raffaghelli, 2022). The EDLSA questionnaire design is a paramount starting point for assessing data literacy competencies in K-12 settings. It offers a flexible framework that other researchers can adapt, test, and refine in diverse educational contexts and with various samples. Thus, future studies, particularly those involving larger sample sizes and cross-cultural and longitudinal research, will be essential for validating the instrument's reliability, confirming its dimensions through Confirmatory Factor Analysis (CFA), and assessing its applicability across diverse educational settings.

5. Conclusions

This study arises from an ongoing gap in the lack of a widely accepted framework tailored to the specific needs of K-12 teachers, compounded by the absence of cross-validated instruments for measuring data literacy in education (Cui et al., 2023; Ghodoosi et al., 2023). As AI and big data technologies continue to shape education, these competencies become increasingly vital for educators in managing and interpreting data (Malik, 2020; J. E. Raffaghelli, 2022).

In response to this need, this work introduces the Educators' Data Literacy Self-Assessment (EDLSA), the first instrument designed for primary and secondary education to address the need for a standardized and widely accepted tool to measure data literacy. Following a review of existing frameworks and expert assessment (Donate et al., 2022), the evolving nature of this field led to the application of Exploratory Factor Analysis (EFA) to develop the tool.

The EDLSA design proposes a cohesive framework and a comprehensive assessment tool for teachers' data literacy across three key dimensions: (1) Comprehensive Educational Analytics, which evaluates the use, transfer, and communication of data to enhance educational practices; (2) Educational Problem-Solving Through Data, which measures teachers' ability to use data to address challenges; and (3) Promoting Meta-Learning Students through Data and Ethical Implications, which explores teachers' effectiveness in fostering meta-learning among students, alongside ethical and legal considerations in data use. Results from the EFA demonstrate strong reliability, internal consistency, and well-distributed samples across the proposed dimensions.

This tool establishes a framework for assessing data literacy in K −12 settings and offers flexibility for future adaptation and testing across various educational contexts (Donate et al., 2024). Thus, the design presented is a critical starting point in delving into relationships between constructs by supporting construct validity in measuring key competencies related to data use, privacy, and decision-making in the classroom in more extensive cross-cultural studies (Donate et al., 2022).

5.1. Limitations of the study and future directions

Building on the preliminary evidence provided by this study, the development of the instrument represents an important step toward measuring data literacy competencies in K-12 education. Although the exploratory nature of this research establishes a foundation for further validation and broader application, contextual factors, such as differences between educational centres, cultural nuances, and levels of innovation in data use (Carretero et al., 2017; Elkordy & Iovinelli, 2021; Vuorikari et al., 2022). Thus, it will need to be addressed in future studies, emphasizing the importance of Confirmatory Factor Analysis (CFA) across diverse contexts (López et al., 2015).

The instrument has undergone expert review, piloting, and an initial structure presentation. Minor adjustments to construct distribution may occur during validation, but the theoretical foundation and overall structure are expected to remain consistent. Administering the questionnaire to larger, representative samples using probabilistic sampling methods will enhance its reliability and allow a more nuanced analysis of trends and gaps across K 12 educational stages.

Furthermore, addressing internal consistency and construct validity through convergent and discriminant methods (Glass & Arnkoff, 1997) will strengthen the instrument's rigour. Considering international disparities in digital literacy perceptions (Gouseti et al., 2023), conducting pilot studies in diverse educational contexts and nations is essential to assess its adaptability and promote transferability.

Future research should also focus on collecting and analyzing teachers' data literacy practices. Insights from such studies could guide the design of personalized professional development programs, enabling educators to address critical data competencies and effectively integrate data-informed strategies into their teaching practices.

CRediT authorship contribution statement

Belén Donate-Beby: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Francisco José García-Peñalvo:** Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Daniel Amo-Filva:** Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Sofía Aguayo-Mauri:** Visualization, Validation, Software, Data curation.

Data availability statement

The data supporting this study's findings are available from the University of Salamanca.The data can be made available upon reasonable request, subject to approval by the the University of Salamanca's Ethical committee and adherence to ethical guidelines.

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

During the preparation of this work, the author(s) used Grammarly to edit the text's language. After using this tool/service, the author(s) reviewed and edited the content as needed and took (s) full responsibility for the publication's content.

Declaration of competing interest

None.

	ID_1_1	ID_1_2	ID_1_3	US1_2_4	US1_2_5	US1_2_6	US1_2_7	US1_2_8	US2_2_9	US2_2_10	US2_2_11	US2_2_12	TR1_3_13
ID_1_1		.796**	.660	.667	.519	.574	.482	.472	.560	.513	.514	.479	.483
ID_1_2			.718**	.647	.644	.683	.572	.597	.647	.677	.566	.578	.505
ID_1_3				.747**	.691	.697	.592	.516	.752**	.720**	.503	.510	.446
US1_2_4					.723**	.690	.591	.561	.622	.634	.595	.591	.523
US1_2_5						.834**	.739**	.617	.679	.671	.613	.636	.533
US1_2_6							.823**	.759**	.741**	.708**	.704**	.673	.569
US1_2_7								.823**	.636	.585	.653	.680	.596
US1_2_8									.707**	.659	.692	.723**	.600
US2_2_9										.895**	.691	.699	.540
US2_2_10											.661	.653	.502
US2_2_11												.924***	.783**
US2_2_12													.768**

	TR1_3_14	TR1_3_15	TR1_3_16	TR2_3_17	TR2_3_18	TR2_3_19	EVA1_4_20	EVA1_4_21	EVA1_4_22	EVA1_4_23	EVA2_4_24	EVA2_4_25	EVA2_4_26
ID_1_1	.439	.419	.480	.331	.280	.274	.447	.430	.413	.420	.519	.459	.446
ID_1_2	.403	.414	.467	.387	.341	.359	.457	.457	.523	.471	.520	.507	.544
ID_1_3	.372	.399	.416	.467	.519	.528	.504	.437	.481	.508	.644	.584	.599
US1_2_4	.463	.561	.541	.531	.591	.535	.530	.547	.574	.564	.596	.547	.599
US1_2_5	.497	.592	.537	.499	.559	.489	.534	.593	.646	.552	.520	.504	.555
US1_2_6	.472	.586	.597	.571	.620	.526	.574	.641	.630	.628	.609	.549	.616
US1_2_7	.531	.651	.634	.577	.603	.533	.585	.650	.673	.644	.599	.607	.621
US1_2_8	.522	.652	.676	.589	.594	.489	.601	.675	.670	.689	.545	.511	.597
US2_2_9	.540	.518	.575	.571	.584	.544	.728**	.603	.640	.677	.646	.622	.658
US2_2_10	.448	.436	.484	.484	.559	.524	.697	.627	.655	.638	.560	.535	.555
US2_2_11	.736**	.802**	.810**	.755**	.691	.627	.800**	.849**	.849**	.848**	.695	.680	.711**
US2_2_12	.755**	.797**	.820**	.747**	.689	.638	.773**	.821**	.843**	.844**	.701**	.685	.721**
TR1_3_13	.846**	.859**	.879**	.679	.653	.658	.778**	.816**	.835**	.839**	.627	.621	.576
TR1_3_14		.794**	.793**	.633	.615	.614	.788**	.762**	.763**	.797**	.558	.595	.556
TR1_3_15			.893**	.734**	.703**	.638	.727**	.793**	.820**	.820**	.614	.601	.631
TR1_3_16				.703**	.626	.555	.766**	.854**	.809**	.845**	.655	.578	.621
TR2_3_17					.875**	.806**	.717**	.702**	.700**	.770**	.846**	.841**	.848**
TR2_3_18						.926***	.693	.693	.691	.745**	.754**	.748**	.745**
TR2_3_19							.663	.603	.655	.693	.690	.732**	.710**
EVA1_4_20								.859**	.802**	.847**	.657	.660	.650
EVA1_4_21									.900***	.835**	.612	.574	.587
EVA1_4_22										.845**	.617	.621	.600
EVA1_4_23											.716**	.698	.672
EVA2_4_24												.930***	.901***
EVA2_4_25													.907***
EVA2_4_26													

Figure Appendix 5.

Acknowledgements

This work has been carried out with funds from Project SPADATAS (Security and Privacy in Academic Data Management at Schools) (Reference number: 2022-1-ES01-KA220-SCH-000086363). The European Union - Next Generation EU, from the Transformation and Resilience Plan and the Public Spanish Employment Service (SEPE), and the Spain Ministry of Science and Innovation (PID2020-118345RB-I00) also provided financial support for this paper.

Data availability

Data will be made available on request.

References

Abdullah, A. (2022). Artificial intelligence in education: A bibliometric analysis (2010–2021). https://www.semanticscholar.org/paper/Artificial-Intelligence-in-Education%3A-A-Analysis-Abdullah/5cee344732c26f3d26adf7a94eac2639be941164.

Alier, M., Pereira, J., García-Peñalvo, F. J., Casañ, M. J., & Cabré, J. (2025). Lamb: An open-source software framework to create artificial intelligence assistants deployed and integrated into learning management systems. *Computer Standards & Interfaces, 92*, Article 103940. https://doi.org/10.1016/j.csi.2024.103940

Alier-Forment, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A., & Llorens-Largo, F. (2024). Safe AI in education Manifesto. Version 0.4.0 https://manifesto.safeaieducation.org.

Amo-Filva, D., Escudero, D. F., Sanchez-Sepulveda, M. V., García-Holgado, A., García-Holgado, L., García-Peñalvo, F. J., Orehovački, T., Krašna, M., Pesek, I., Marchetti, E., Valente, A., Witfelt, C., Ružić, I., Fraoua, K. E., & Moreira, F. (2023). Security and privacy in academic data management at schools: SPADATAS Project. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies* (Vol. 14040, pp. 3–16). Nature Switzerland: Springer. https://doi.org/10.1007/978-3-031-34411-4_1.

Amo-Filva, D., Fonseca, D., García-Peñalvo, F. J., Forment, M. A., & Guerrero, M. J. C. (2023). Learning analytics' privacy in the fog and edge computing: A systematic mapping review. In F. J. García-Peñalvo, & A. García-Holgado (Eds.), *Proceedings TEEM 2022: Tenth international conference on technological ecosystems for enhancing multiculturality* (pp. 1199–1207). Singapore: Springer Nature. https://doi.org/10.1007/978-981-99-0942-1_126.

Amo-Filva, D., Prinsloo, P., Alier Forment, M., Fonseca Escudero, D., Torres Kompen, R., Canaleta Llampallas, X., & Herrero Martín, J. (2021). Local technology to enhance data privacy and security in educational technology. *International Journal of Interactive Multimedia and Artificial Intelligence, 7*(2), 262–273.

Amo-Filva, D. A., Campión, R. S., & Prieto, A. (2017). *Learning analytics: La narración del aprendizaje a través de los datos*. Editorial UOC. https://dialnet.unirioja.es/servlet/libro?codigo=746325.

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). *Self-efficacy: The exercise of control*. Springer. https://connect.springerpub.com/content/sgrjcp/13/2/158.full.pdf.

Biezā, K. E. (2020). Digital literacy: Concept and definition. *International Journal of Smart Education and Urban Society (IJSEUS), 11*(2), 1–15. https://www.igi-global.com/article/digital-literacy/248468.

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology, 9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://www.emerald.com/insight/content/doi/10.3316/QRJ0902027/full/html.

Bowers, A. J. (2017). Quantitative research methods training in education leadership and administration preparation programs as disciplined inquiry for building school improvement capacity. *Journal of Research on Leadership Education, 12*(1), 72–96. https://doi.org/10.1177/1942775116659462

Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450. https://psycnet.apa.org/record/2016-15750-001.

Brandl, R., & Ellis, C. (2023). ChatGPT statistics: A deep dive into the numbers behind ChatGPT. *Tooltester*. https://www.tooltester.com/en/blog/chatgpt-statistics/.

Buckingham, D. (2010). In B. Bachmair (Ed.), *Defining digital literacy* (pp. 59–71). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92133-4_4.

Buckingham Shum, S. (2012). Learning analytics. *Society for Learning Analytics Research (SoLAR)*. https://www.solaresearch.org/core/unesco-policy-brief-learning-analytics/.

Burke, A., & Mann, A. (2024). *Digital technologies in career guidance for youth: Opportunities and challenges*. OECD Publishing. https://doi.org/10.1787/c9ab23da-en

Caena, F., & Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (Digcompedu). *European Journal of Education, 54*(3), 356–369. https://doi.org/10.1111/ejed.12345

Carretero, S., Vuorikari, R., & Punie, Y. (2017). DigComp 2.1: The digital competence framework for citizens. https://apo.org.au/node/221736.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

COMEST. (2019). *Preliminary study on the ethics of artificial intelligence*. Paris, France: UNESCO. https://ircai.org/wp-content/uploads/2020/07/preliminary-study-on-the-ethics-of-artificial-intelligence.pdf.

Creswell, J. W. (2012). Educational research. *Planning, conducting, and evaluating* (4th ed.). Pearson.

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. https://doi.org/10.1080/0969594X.2023.2182737

de Yébenes Prous, M. J. G., Salvanés, F. R., & Ortells, L. C. (2009). Validation of questionnaires. *Reumatología Clínica, 5*(4), 171–177. https://www.sciencedirect.com/science/article/pii/S2173574309701157.

Donate, B., García-Peñalvo, F. J., Amo, D., De Torres, E., & Herrero-Martín, J. (2022). An instrument for self-assessment of data literacy at the micro, Meso, and macro educational levels. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies. Designing the learner and teacher experience* (Vol. 13328, pp. 228–237). Springer International Publishing. https://doi.org/10.1007/978-3-031-05657-4_16.

Donate, B., García-Peñalvo, F., & Amo-Filva, D. (2024). *Data literacy questionnaire for educators*. https://doi.org/10.5281/zenodo.10453796

Drouin, M., McDaniel, B. T., Pater, J., & Toscos, T. (2020). How parents and their children used social media and technology at the beginning of the COVID-19 pandemic and associations with anxiety. *Cyberpsychology, Behavior, and Social Networking, 23*(11), 727–736. https://doi.org/10.1089/cyber.2020.0284

Elkordy, A., & Iovinelli, J. (2021). Competencies, culture, and change: A model for digital transformation in K-12 educational contexts. In *Digital transformation of learning organizations* (Vol. 203). https://library.oapen.org/bitstream/handle/20.500.12657/46121/1/2021_Book_DigitalTransformationOfLearnin.pdf#page=217.

Feldt, L. S., & Charter, R. A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological Methods, 8*(1), 102. https://psycnet.apa.org/journals/met/8/1/102/.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304. https://doi.org/10.1504/IJTEL.2012.051816

Flores-Vivar, J. M. F., & García-Peñalvo, F. J. G. (2023). Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4). *Comunicar, 31*(74), 35–44. https://doi.org/10.3916/C74-2023-03

Frías-Navarro, D., & Pascual Soler, M. (2012). Prácticas del Análisis Factorial Exploratorio (AFE) en la investigación sobre conducta del consumidory marketing. *Suma Psicológica, 19*(1), 47–58. http://www.scielo.org.co/scielo.php?pid=s0121-43812012000100004&script=sci_arttext.

García-Peñalvo, F. J., & Corell, A. (2020). La COVID-19: ¿enzima de la transformación digital de la docencia o reflejo de una crisis metodológica y competencial en la educación superior? *Campus Virtuales, 9*(2), 83–98.

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780429056765/ibm-spss-statistics-26-step-step-darren-george-paul-mallery.

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

Glass, C. R., & Arnkoff, D. B. (1997). Questionnaire methods of cognitive self-statement assessment. *Journal of Consulting and Clinical Psychology, 65*(6), 911. https://psycnet.apa.org/record/1998-00068-002.

Gouseti, A., Lakkala, M., Raffaghelli, J., Ranieri, M., Roffi, A., & Ilomäki, L. (2023). Exploring teachers' perceptions of critical digital literacies and how these are manifested in their teaching practices. *Educational Review*, 1–35. https://doi.org/10.1080/00131911.2022.2159933

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record: The Voice of Scholarship in Education, 117*(4), 1–22. https://doi.org/10.1177/016146811511700401

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Hazzi, O., & Maldaon, I. (2015). A pilot study: Vital methodological issues. *Business: Theory and Practice, 16*(1), 53–62. https://jau.vgtu.lt/index.php/BTP/article/view/8305.

Hegestedt, R., Nouri, J., Rundquist, R., & Fors, U. (2023). Data-driven school improvement and data-literacy in K-12: Findings from a Swedish national program. *International Journal of Emerging Technologies in Learning (Online), 18*(15), 189. https://search.proquest.com/openview/643d664d9e32128ccc6a59237755f6cf/1?pq-origsite=gscholar&cbl=5452619&casa_token=w2z2Tlqh7cQAAAAA:U3xuxeTosnbPt5yVU9moRl2tpoUgrxHHzY1WA21f_EOj0dgAis1jIx94mN2D36sBJmIX4w_3.

Henderson, J., & Corry, M. (2021). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning, 14*(2), 232–244. https://www.emerald.com/insight/content/doi/10.1108/JRIT-11-2019-0074/full/html.

Julien, H. (2018). Digital literacy in theory and practice. In *Encyclopedia of information science and technology* (4th ed., pp. 2243–2252). IGI Global https://www.igi-global.com/chapter/digital-literacy-in-theory-and-practice/183937.

Kılıç, S. (1970). Cronbach s alpha reliability coefficient. *Psychiatry and Behavioral Sciences, 6*(1), 47–47 https://www.pbsciences.org/index.php?mno=221091.

Kwak, S. (2023). Are only p-values less than 0.05 significant? A p-value greater than 0.05 is also significant. *Journal of Lipid and Atherosclerosis, 12*(2), 89. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10232224/.

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy" – an exploratory factor

analysis. *Computers in Human Behavior Reports, 12*, Article 100338. https://doi.org/10.1016/j.chbr.2023.100338
Law, N., Woo, D., Torre, J. de la, & Wong, G. (2018). A global framework of reference on digital literacy skills for indicator 4.4. 2. https://policycommons.net/artifacts/8218331/a-global-framework-of-reference-on-digital-literacy-skills-for-indicator-442/9132347/.
Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research, 37*(4), 412–421. https://doi.org/10.1080/01488376.2011.580697
Lisitsa, E., Benjamin, K. S., Chun, S. K., Skalisky, J., Hammond, L. E., & Mezulis, A. H. (2020). Loneliness among young adults during covid-19 pandemic: The mediational roles of social media use and social support seeking. *Journal of Social and Clinical Psychology, 39*(8), 708–726. https://doi.org/10.1521/jscp.2020.39.8.708
Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3313831.3376727
Long, P., & Siemens, G. (2011). What is learning analytics. In *Proceedings of the 1st international conference learning analytics and knowledge* (Vol. 11)LAK. https://solaresearch.org/wp-content/uploads/hla22/HLA22_Chapter_1_Lang.pdf.
López, X., Valenzuela, J., Nussbaum, M., & Tsai, C.-C. (2015). Some recommendations for the reporting of quantitative studies. https://psycnet.apa.org/record/2015-53292-010.
Malik, M. M. (2020). A hierarchy of limitations in machine learning. https://doi.org/10.48550/arXiv.2002.05193.
Mandinach, E. B., & Abrams, L. M. (2022). Data literacy and learning analytics. *Handbook of Learning Analytics*, 196–204. https://solaresearch.org/wp-content/uploads/hla22/HLA22_Chapter_19_Mandinach.pdf.
Mandinach, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education, 60*, 452–457.
Markham, A. N., Tiidenberg, K., & Herman, A. (2018). Ethics as methods: Doing ethics in the era of big data research—introduction. *Social Media + Society, 4*(3), Article 2056305118784502. https://doi.org/10.1177/2056305118784502
Maybee, C., & Zilinski, L. (2015). *Data informed learning: A next phase data literacy framework for higher education*.
Melnikova, J., Batuchina, A., Ahrens, A., & Zascerinska, J. (2023). *Teachers' data literacy skills for pedagogical decision making: Needs analysis in Lithuania and Germany* (Vol. 2, pp. 182–188). https://doi.org/10.17770/etr2023vol2.7287. Scopus.
Naamati-Schneider, L., & Alt, D. (2024). Beyond digital literacy: The era of AI-powered assistants and evolving user skills. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12694-z
Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education, 14* (3). https://sdme.kmu.ac.ir/article_90598.html.
Olari, V., & Romeike, R. (2021). Addressing AI and data literacy in teacher education: A review of existing educational frameworks. https://computingeducation.de/pub/2021_Olari-Romeike_WIPSCE21.pdf.
O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. https://docs.metea.org.uk/lib/MFTFTRFL.
Oviedo, H. C., & Campo-Arias, A. (2005). An approach to the use of Cronbach's Alfa. *Revista Colombiana de Psiquiatria, 34*(4), 572–580. http://www.scielo.org.co/scielo.php?pid=S0034-74502005000400009&script=sci_abstract&tlng=pt.
Peñalvo, F. J. G., Alier, M., Pereira, J., & Casany, M. J. (2024). Safe, transparent, and ethical artificial intelligence: Keys to quality sustainable education (SDG4). *IJERI: International Journal of Engineering Research & Innovation, 22*, 1–21. https://www.upo.es/revistas/index.php/IJERI/article/view/11036.
PISA. (2003). The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills. http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf.
Raffaghelli, J. E. (2018). Educators' data literacy. *MEDIA EDUCATION: STUDI, RICERCHE, BUONE PRATICHE, teacher education&Training on ICT between Europe and Latin America*, 91–109 https://doi.org/10.4399/97888255210238.
Raffaghelli, J. (2019). «Datificación» y Educación Superior: Hacia la construcción de un marco para la alfabetización en datos del profesorado universitario. https://doi.org/10.15332/25005421/5466.
Raffaghelli, J. E. (2022). Educators' data literacy: Understanding the bigger picture. In *Learning to live with datafication* (pp. 80–99). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003136842-5/educators-data-literacy-juliana-raffaghelli.
Redecker, C., & Punie, Y. (2017). In Y. Punie (Ed.), *Digital competence of educators*. https://core.ac.uk/download/pdf/132627227.pdf.
Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D. R., Matwin, S., & Wuetherick, B. (2015). Strategies and best practices for data literacy education: Knowledge synthesis report. https://doi.org/10.13140/RG.2.1.1922.5044.
Rosenthal, R. (1968). An application of the Kolmogorov-Smirnov test for normality with estimated mean and variance. *Psychological Reports*. https://psycnet.apa.org/record/1968-11379-001.
Salomão Filho, A., Wasson, B., Lameras, P., & Maloszek, R. (2023). Data literacy for citizenry: A few policy recommendations from a literature review. *Edutec. Revista Electrónica de Tecnología Educativa, 86*, 6–21. https://pureportal.coventry.ac.uk/en/publications/data-literacy-for-citizenry-a-few-policy-recommendations-from-a-l.
Saris, W. E., De Pijper, M., & Mulder, J. (1978). Optimal procedures for estimation of factor scores. *Sociological Methods & Research, 7*(1), 85–106. https://doi.org/10.1177/004912417800700104
Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports, 1*, Article 100014. https://doi.org/10.1016/j.chbr.2020.100014
Selwyn, N., Pangrazio, L., Nemorin, S., & Perrotta, C. (2020). What might the school of 2030 be like? An exercise in social science fiction. *Learning, Media and Technology, 45* (1), 90–106. https://doi.org/10.1080/17439884.2020.1694944
Sharma, A., Gupta, A., Bhargava, A., Rawat, A., Yadav, P., & Gupta, D. (2023). From sci-fi to reality: The evolution of human-computer interaction with artificial intelligence. *2023 2nd international conference on applied artificial intelligence and computing (ICAAIC)*. https://ieeexplore.ieee.org/abstract/document/10141431/?casa_token=pE_CjKE-cOYAAAAA:g1LS9P_25TvSpw81YUts1mNlWZFGggYq2eWPCZvzSLhImuRMMM2QADZgQ3BJnN9RpTSbOGvH.
Shields, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly, 28*(2–3), 6–6 https://iassistquarterly.com/index.php/iassist/article/view/790/782.
Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851
Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd international conference on learning analytics and knowledge*. https://doi.org/10.1145/2330601.2330661
Snyder, L. G., & Snyder, M. J. (2008). Teaching critical thinking and problem solving skills. *The Journal of Research in Business Education, 50*(2), 90. http://dme.childrenshospital.org/wp-content/uploads/2019/02/Optional-_Teaching-Critical-Thinking-and-Problem-Solving-Skills.pdf.
Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1–A9. Elsevier https://www.sciencedirect.com/science/article/pii/S0360131516302445?casa_token=rwIZITywrroAAAAA:aAlA-WKYnPoMuZozcqZCGfJpjowxN3hZ_cQDO9VsaJcUbu37G1myw94yFIEJRFiSgL4JE6Mh.
UNESCO. (2022). K-12 AI curricula: A mapping of government-endorsed AI curricula (ED-2022/FLI-ICT/K-12 REV) https://bit.ly/41Jtsi8.
Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens - with new examples of knowledge, skills and attitudes. *JRC Publications Repository*. https://doi.org/10.2760/115376
Wadkar, S. K., Singh, K., Chakravarty, R., & Argade, S. D. (2016). Assessing the reliability of attitude scale by Cronbach's alpha. *Journal of Global Communication, 9*(2), 113–117. https://www.indianjournals.com/ijor.aspx?target=ijor:jgc&volume=9&issue=2&article=008.
Wang, Z. (2020). Exploring different notions of literacy: A literature review analysis of literacy research related to artificial intelligence and big data application. *IOP Conference Series: Materials Science and Engineering, 806*(1), Article 012023. https://iopscience.iop.org/article/10.1088/1757-899X/806/1/012023/meta.
White, B. Y., Shimoda, T. A., & Frederiksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: Computer support for metacognitive development. *International Journal of Artificial Intelligence in Education, 10*, 151–182. https://telearn.hal.science/hal-00197340/.
Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of Big Data. *Journal of Learning Analytics, 2*(2), 5–13.
Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics, 12*(3), 3. https://doi.org/10.15353/joci.v12i3.3275
Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies, 129*, 41–54. https://doi.org/10.1016/j.ijhcs.2019.03.006
Zaytseva, T., Kravtsova, L., & Puliaieva, A. (2019). Computer modelling of educational process as the way to modern learning technologies. http://rep.ksma.ks.ua/handle/123456789/440.