JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 12(2), 175–195. https://doi.org/10.18608/jla.2025.8441

Assessing Data Landscapes for Quality Education in Latin America: A FAIRness Perspective on Brazil, Colombia, and Peru

Rosa R. Soto Ruidias[1*], Bernardo Pereira Nunes[2], Ruben Manrique[3] and Sean Siqueira[4]

Abstract

Despite the increasing availability of data used to inform educational policies and practices, concerns persist regarding its quality and accessibility. This study surveys quality education data from Brazil, Colombia, and Peru and evaluates their alignment with the FAIR principles and availability to support academic analytics (AA) and learning analytics (LA). We identified and analyzed 112 data sources, from which 93% of the data sets originate from government repositories and open data platforms, with coverage of macro-level data relevant for AA but lack of granularity for LA. The FAIR assessment showed 50% of compliance with findability (F), 33% for accessibility (A), and less than 50% for both interoperability (I) and reusability (R), which limits broader utility. Moreover, these diverse data sources present limitations in quality assurance metrics such as "institutional development" and "quality management." We conclude by offering recommendations, emphasizing the need for enhanced data frameworks that bridge macro- and micro-level data for AA and LA to enable data-driven decisions for improving educational quality in Latin America.

Notes for Practice

- This study evaluates the quality, availability, and compliance with FAIR principles of existing quality education data in Brazil, Colombia, and Peru, emphasizing their importance for educational decision-making in Latin America.
- Current data sets support academic analytics (AA) but lack the granular behavioural, learner disposition, and biometric data needed for learning analytics (LA), highlighting the limitations of observatories and open data sets in this area.
- The study uncovers nuances for educational improvement, identifying data issues through FAIR principles and offering actionable recommendations to advance educational practices and policies.
- The paper underscores the need for policymakers to prioritize data quality and accessibility, advocating for policies that promote transparency and standardization in data collection and sharing practices.

Keywords

Quality in education, educational data observatories, FAIR principles, education in Latin America.

Submitted: 10/04/2024 — **Accepted:** 11/06/2025 — **Published:** 19/08/2025

[1*] *Corresponding author email: rosarosmery.sotoruidias@anu.edu.au Address: School of Computing, College of Systems & Society, Australian National University, Canberra, Australia. ORCID iD: https://orcid.org/0009-0005-3522-3807*

[2] *Email: bernardo.nunes@anu.edu.au Address: School of Computing, College of Systems & Society, Australian National University, Canberra, Australia. ORCID iD: https://orcid.org/0000-0001-9764-9401*

[3] *Email: rf.manrique@uniandes.edu.co Address: Universidad de los Andes, Bogota, Colombia. ORCID iD: https://orcid.org/0000-0001-8742-2094*

[4] *Email: sean@uniriotec.br Address: Federal University of the State of Rio de Janeiro, Rio de Janeiro, Brazil. ORCID iD: https://orcid.org/0000-0002-0864-2396*

1. Introduction

Quality education is a foundation of social and economic development, playing a fundamental role in shaping the future of nations and individuals. In Latin America (LATAM), efforts to improve educational quality have made significant progress (UNESCO, 2023), but challenges remain. Addressing these challenges requires a well-structured and comprehensive approach to understanding and enhancing education systems (OECD, 2024), emphasizing the effective use of educational data to inform policy, ensure accountability, and drive continuous improvement. This research is part of the Educational Data Observatory Network project designed to bridge the divide between available data and the wider insights needed to inform educational policies and practices.

In this study, the data landscape for quality education is defined as the state of public data sets, platforms, and observatories related to quality assurance in education and its applicability for learning analytics (LA), which examines micro-level learning dynamics, and academic analytics (AA), which focuses on broader, macro-level trends (OECD, 2021b; Nguyen et al., 2020). Moreover, we examine issues drawn from the FAIR data framework (Wilkinson et al., 2016). By integrating LA and AA, institutions can gain better comprehension of student success, identify areas for improvement, and make data-driven decisions to enhance the learning experience (Nguyen et al., 2020).

The FAIR principles, guiding data to be findable (F), accessible (A), interoperable (I), and reusable (R), promote broader sharing and reuse of data (FAIRsFAIR, 2019). This framework enhances data stewardship, fosters collaboration, and facilitates evidence-driven decision-making (Rivas & Sanchez, 2022), playing a key role in improving educational quality.

Table 1 provides the characteristics of the educational systems in Brazil, Colombia, and Peru, which, despite their distinct approaches, all adhere to traditional educational levels (primary, secondary, tertiary, and technical). Brazil uses various evaluations and examinations, such as the National Assessment of Basic Education (ANEB) and the National High School Exam (ENEM), emphasizing authorization and accreditation. Colombia employs the Synthetic Index of Educational Quality (ISCE) and the National Accreditation System (CNA) for rigorous quality standards. In Peru, quality control is managed by the National Policy for Higher and Technical-Productive Education (PNESTP), the National Superintendency of Higher University Education (SUNEDU), and the National System of Evaluation, Accreditation, and Certification of Educational Quality (SINEACE), with assessments like ECE and PISA. According to World Bank (2022), government expenditure on education as a percentage of GDP was 4.2% for Peru in 2023, 5.5% for Brazil in 2021, and 3.7% for Colombia in 2001.

Table 1. Characteristics of the educational systems in Brazil, Colombia, and Peru.

Education System Structure		
Brazil (Jamil Cury, 2022; OECD, 2021a; "Ministério da Educação", n.d.; Fujikawa, 2015)	**Colombia** (CNA, 2022; Ministerio de Educación Nacional, 2021; OECD, 2023; ICFES, 2012)	**Peru** (Congreso de la República, n.d., p. 28044); (SUNEDU, 2015; UNESCO, 2017)
Divided into early childhood, primary (mandatory), secondary, and tertiary education (university and technology).	Includes primary education, compulsory basic secondary education (grades 6 to 9), optional middle vocational education (grades 10 and 11), and tertiary education with technical institutes, technological institutions, and academic universities.	Comprises basic education (regular, alternative, and special), technical-productive education, and higher education (undergraduate and postgraduate university, pedagogical, technological, and artistic training).
Quality Policy Focus and Assessment Mechanisms		
Emphasis on authorization, recognition, accreditation, and various assessments to ensure quality, such as the Index of Development of Basic Education—IDEB.	Focus on the ISCE and multiple assessments at different stages.	National Policy for Higher and Technical-Productive Education (PNESTP) with SUNEDU oversight, ensuring quality control and promotion.
Various evaluations and examinations at different educational levels. Includes ANEB, ENEM, IDEB, Prova Brasil, SEEDUC, PISA, and ENADE for higher education.	Uses the ISCE, focusing on performance, progress, efficiency, and school environment. Employs assessments like SABER, ECAES, PISA, and ERCE.	Quality control through the PNESTP with SUNEDU oversight. Assessments include ECE, LLECE, PISA, and ICCS.
Evaluate, supervise, and coordinate quality accreditation in higher education with the National Higher Education Evaluation System (SINAES) and the Commission (CONAES).	CNA employs a mixed model for accreditation, ensuring rigorous quality standards.	Accreditation overseen by SINEACE.

The existing literature lacks studies on public educational data and their alignment with FAIR principles. Studies frequently overlook evaluating data sets themselves, including their adherence to FAIR principles and their suitability for enhancing quality education assessments (Inusah et al., 2022; Mihaescu & Popescu, 2021; Kumar Verma & Shrivastava, 2015). While public educational data observatories (EDOs) offer valuable insights into education systems and support accountability and policy-making, they are often designed for federal reporting rather than fostering collaboration or broad accessibility for stakeholders in LA and AA (Mandinach & Jimerson, 2022; Mandinach & Schildkamp, 2021). Moreover, another gap is the inconsistency in national and global definitions of quality education indicators, which interfere with the interoperability and consistent use of data (UNESCO, 2023). Despite efforts by international organizations and local governments in data collection and quality assessment, there remains a need to enhance the accessibility, transparency, and interoperability of educational data to better support LA and AA (Glick et al., 2019; Proética, 2020).

This study investigates the state of educational data in Brazil, Colombia, and Peru, specifically focusing on data related to quality assurance of education. Our primary objective is to assess the alignment of these data with the FAIR principles and to identify key challenges and opportunities for leveraging data to support AA and LA and enable cross-country comparisons. We address the following research question: "How do existing EDOs and data sets in these countries align with FAIR principles, and what are the limitations in data quality, granularity, and coverage for supporting LA and AA?"

2. Literature Review

The increasing availability of educational data contained within EDOs is fundamental for improving teaching and learning (Shanks et al., 2018; Somogyvári et al., 2023). International organizations (UNESCO, OECD, World Bank, UNDP) and national ministries (Peru's MINEDU, Brazil's INEP, Colombia's MEN) (Eniceia & Fabiana, 2015; Mariano et al., 2022; van Oostveen et al., 2019) promote such platforms, offering data from student enrolment to teacher qualifications (Scavarda et al., 2023). While AA has benefited from increased data accessibility (OECD, 2021b), LA faces challenges accessing data beyond institutional restrictive networks (Glick et al., 2019). Limitations in data extraction and integration (Varella Ehrenfried et al., 2019; A. Bowers et al., 2022), coupled with insufficient evaluation of data and repositories (Inusah et al., 2022; Mihaescu & Popescu, 2021; Kumar Verma & Shrivastava, 2015), necessitate evaluating and advocating for transparency and robust data management (Proética, 2020). Furthermore, the diverse quality assessment instruments used in Peru, Brazil, and Colombia highlight the need for a holistic perspective (Karakhanyan & Stensaker, 2020).

2.1 EDOs

EDOs play a critical role in monitoring and promoting the quality of education. Government agencies, educational institutions, and international organizations are the main creators of these repositories to collect, analyze, and disseminate education-related data (Shanks et al., 2018). These platforms offer valuable insights into education systems (Somogyvári et al., 2023). For example, the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Organisation for Economic Co-operation and Development (OECD), the World Bank, and the United Nations Development Programme (UNDP) offer multiple observatories that provide evidence-based information to inform policies, programs, and initiatives aimed at enhancing education and driving sustainable development worldwide.

In Peru, Brazil, and Colombia, organizations such as the Peruvian Ministry of Education (MINEDU), the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), and the Colombian Ministry of National Education (MEN), respectively, and other government research institutions have developed data observatories (Eniceia & Fabiana, 2015; Mariano et al., 2022) to consolidate data from different sources (van Oostveen et al., 2019). These observatories encompass a wide range of data, including student enrolment, learning outcomes, teacher qualifications, and infrastructure statistics (Scavarda et al., 2023), analyzed by AA. Because AA has a general scope of data requirement, its accessibility has been covered by the regulation, so institutions, government, and international educational agencies are managing to improve the accessibility of the data (OECD, 2021b). However, Glick and colleagues (2019) emphasize that publications about LA have broad access to diverse student data and learning interactions only at the institutional level, but that limited access to educational data outside institutional networks may disadvantage stakeholders.

While EDOs provide valuable information, their content and coverage need examination. Prior research highlights limitations in data extraction, formatting, and integration from repositories and OpenData platforms (Varella Ehrenfried et al., 2019; A. Bowers et al., 2022). Other studies focus on data mining without thoroughly evaluating the data sets, repositories, or observatories they use (Inusah et al., 2022; Mihaescu & Popescu, 2021; Kumar Verma & Shrivastava, 2015). Evaluating these aspects is crucial for ensuring the quality, fairness, and transparency of findings; addressing biases; adhering to ethical standards; and enhancing reliability and applicability. For instance, Proética (2020) emphasizes scrutinizing higher educational data and advocating for greater transparency, public access, and constant updates.

2.2 Quality Education Assessment in LATAM

Brazil, Colombia, and Peru have similar education quality assessment instruments. In Brazil, the SINAES assesses institutions, courses, student performance, teaching, research, extension, social responsibility, student performance, institutional management, faculty, and facilities. CNA in Colombia and SINEACE and SUNEDU in Peru are equivalent institutions.

UNESCO assesses high quality and equity within the framework of Sustainable Development Goal 4 (SDG) (Murillo & Cuenca, 2007), while the OECD considers education attainment, education resources, PISA, enrolment, graduation, student mobility, teacher salaries, student dropout rates, teacher ages, and youth in the labour market. The different government regulations (internal to the country) and approaches by international organizations (external to the country) present a wide variety of indicators that are consolidated in guidelines such as the Model for Quality Assurance in University Education (SUNEDU, 2021).

The studies presented in the special section on applications of LA (Hilliger et al., 2024) in LATAM are primarily concentrated on adopting and putting into practice LA solutions to enhance the quality of education, in the same line as the research done in Europe and Australia (Siemens & Long, 2011; European Commission. Joint Research Centre, 2016). This special section also points out that the development and adoption of LA in LATAM is limited by insufficient high-quality learning data and the lack of capacity to effectively collect and analyze the data. Among the ongoing research in the region, Talamás-Carvajal and colleagues (2024) highlight that developing critical thinking significantly contributes to the advancement of complex thinking skills and related sub-competencies, particularly scientific and systemic thinking. Consequently, it is recommended that curricula be reformulated to support the development of critical thinking from the early semesters. This underscores how access to LA data can inform decision-making in teaching and learning centres. However, the study acknowledges a key limitation: the data were obtained from a single institution. To improve generalizability, future research should incorporate data from multiple institutions. Therefore, ensuring that data are FAIR is essential to facilitate access and enable further advances in LA for decision-making.

Additionally, Bautista Godínez and colleagues (2024) emphasize growing interest in using data for feedback, profiling, and understanding the link between performance and well-being, with themes like health, relationships, and governance informing future LA initiatives. Heterogeneous sample sizes and limited access to detailed sociodemographic data, among others, represent the limitations of the study. These challenges underscore the importance of implementing FAIR data practices to enable more robust and insightful LA research. Furthermore, MMALA (Freitas et al., 2024) and PROF-XXI (Kotorov et al., 2024) are other regional efforts to strengthen institutional capacities for using educational data effectively, emphasizing the importance of structured, context-sensitive approaches to support quality and innovation in higher education institutions. AA and LA have increasingly become integral to the assessment and enhancement of education as they help to track student performance, identify at-risk students, and implement targeted interventions to improve educational outcomes.

A UNESCO report (UNESCO, 2023) underscores critical gaps in data availability for government policy decisions, particularly in education quality assessment. They are (1) absence of relevant data, (2) discrepancies between national and global definitions, and (3) insufficient communication and collaboration resources. It also highlights the efforts of organizations to improve data availability, emphasizing the urgent need for enhanced data collection, especially in learning assessment. This context demands our study, which aims to assess the existing data landscapes in Brazil, Colombia, and Peru, with a particular emphasis on evaluating data quality and adherence to FAIR principles and identifying the types of data used for AA and LA.

2.3 FAIR Data Management Principles

The FAIR principles (Wilkinson et al., 2016) require comprehensive metadata, open access policies, harmonized data formats, and thorough documentation to facilitate access, integration, and optimization for various stakeholders (FAIRsFAIR, 2019). In education, findability (F) involves detailed metadata and effective indexing, while accessibility (A) includes open access policies and user-friendly portals. Interoperability (I) focuses on harmonizing data formats, enabling integration from different sources. Reusability (R) promotes proper documentation and citation, simplifying the extension and verification of educational research (A. Bowers et al., 2022).

The FAIR principles have gained prominence in data management and accessibility across various fields, including education (N. Khan et al., 2022). The non-educational observatories show multifaceted challenges for comprehensive approaches to ensure that data is FAIR across various domains (Martínez et al., 2023; Lunga et al., 2023; Riginos et al., 2020). These challenges include managing large, heterogeneous data sets; facilitating effective data-driven environmental research; and integrating existing infrastructures into open science frameworks. Applying FAIR principles to an LA repository enhances data reusability through standardized metadata, vocabulary, and data formats, providing significant advantages to researchers seeking relevant data for their inquiries in LA (Wolff et al., 2021). However, educational data management has often received less attention in discussions about the data life cycle than other components (A. J. Bowers & Choi, 2023).

Even though EDOs have started to incorporate FAIR data practices into their operations to ensure that data is not only collected but also maximally leveraged for the benefit of the education community (Sales et al., 2020; UNESCO, 2021), a gap points out that the infrastructures were primarily created for accountability and federal reporting rather than facilitating data sharing and collaboration among community members, researchers, and practitioners (Mandinach & Jimerson, 2022; Mandinach & Schildkamp, 2021). The active promotion of the FAIR implementation calls for improving the management of educational data to maximize its potential impact on decision-making and research.

3. Methodology

This methodology assesses the alignment of educational data with our taxonomy for quality educational assessment (SUNEDU, 2021), FAIR principles, and availability to support AA and LA. The process involved identifying, downloading, and systematically analyzing data sets from publicly accessible national and international sources. Figure 1 summarizes in a flowchart the methodology used in this study. Previous studies (A. Bowers et al., 2022; B. H. Khan et al., 2018; Bloom-Weltman et al., 2018) have emphasized using data science to analyze educational data and create profiles. Key priorities in the assessment process include data privacy, quality, and seamless interoperability between sources, alongside ensuring accessibility in diverse contexts (A. J. Bowers et al., 2019). The FAIR principles for data management and stewardship, as highlighted in A. J. Bowers and Choi (2023), comprehensively address these attributes within the EDO framework, making them relevant to this study.

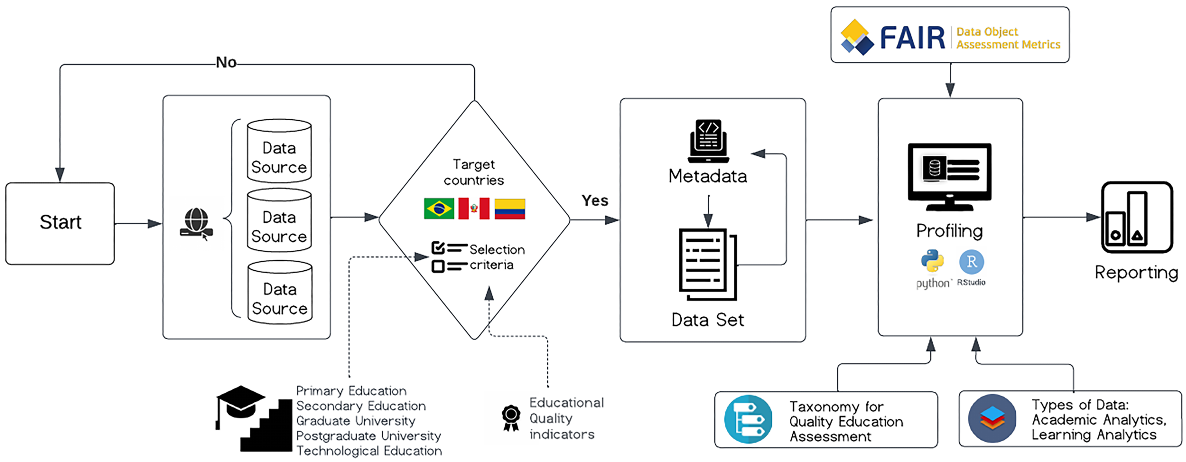

Figure 1. Flowchart of the methodology.

LA focuses on the micro-level, analyzing individual student behaviours, interactions, and learning outcomes; in contrast, AA targets the macro-level, addressing institutional performance, policy development, and system-wide trends to inform strategic decision-making and policy development. "LA is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Siemens & Long, 2011). On the other hand, AA focuses on using institutional data to inform policy and management decisions, analyzing enrolment patterns, retention rates, and financial data to optimize resource allocation and support institutional planning (European Commission. Joint Research Centre, 2016). Consequently, LA also informs AA by providing granular data that can reveal systemic issues and guide broader policy and management decisions. Using the classification by Littlejohn (2022), we identified LA-relevant data types, including behavioural data (e.g., site access frequency), discourse data (e.g., learner communication), learner disposition data (e.g., learning preferences), and biometric data. The distinction between LA and AA, based on focus, data sources, granularity, and purpose, ensures a structured approach to data analysis and interpretation.

To evaluate the utility of data in supporting LA and AA, we established a relationship between the taxonomy metrics and these domains (Section 3.1.3, Table 2). This was achieved by manually classifying each taxonomy metric as relevant to AA, LA, or both, based on the definitions of AA and LA and the operational descriptions of the metrics provided in SUNEDU (2021). Additionally, our Profiling Tool for Data Observatories and Data Sets (Table 3) encompasses various aspects, including intrinsic metadata about a data set, contextual metadata within the data set, the type of measure in relation to the taxonomy, and AA and LA classification, providing the definitions for each element considered for the analysis. The next subsections provide further explanation.

3.1 Criteria for Searching and Including Observatories and Data Sets in the Study

3.1.1 Justification for Selecting Brazil, Colombia, and Peru as Target Countries

Brazil, Colombia, and Peru represent some of the largest education systems in the region. They have substantially implemented policies that reshaped their education systems, including performance-based teacher career regulations, curricular standards enforced through textbooks, high-stakes testing for accountability, and the creation of "school units" for decision-making (Rivas & Sanchez, 2022). Collectively, these three countries constitute 68% of the population in South America and 46% in LATAM and account for more than 62% and 40% of their GDP, respectively (World Bank, 2022).

3.1.2 Criteria for Searching and Including Data Education Observatories and Data Sets

The search for observatories and data sets was conducted by three researchers, each focusing on one of the target countries—Brazil, Peru, and Colombia. These researchers, experts in computer science and education research and native to their respective countries, systematically identified and collected information on observatories, platforms, and repositories that might contain relevant educational data. Based on the formulation of the research question, the keywords (educational data observatories, data repositories) and their associated terms (open data, data hubs, platforms, dashboard, metrics) and their translation into Spanish, Portuguese, and English, were identified and utilized for the search. The collected information included country/region, government/non-government, name, main URL, description, and other relevant URL (dashboards, repositories etc.). Each researcher performed both national and international searches, ensuring that the scope remained aligned with the context and needs of their respective countries.

The criteria require observatories and data sets to be open data, containing quality education indicators for primary, secondary, and tertiary education. They should include a diverse range of educational quantitative data.

3.1.3 Relevance to Quality Education Assessment for the Sample Countries

In order to analyze the content of the observatories and data sets, concerning the assessment of education quality, we have taken a taxonomy based on SUNEDU (2021). This document presents an International Systematization of Quality Assurance Models (it compiles data in a taxonomy that facilitates grouping indicators). In developing this systematization, the authors referenced each country's governmental guidelines, models, or normative standards for quality education internally. Externally, they looked to key global organizations, including the OECD, UNESCO, and the World Bank, renowned for their expertise in the global education landscape and in research (Zapp, 2021).

In Table 2, we present the taxonomy, which is incorporated in Table 3; detailed information about this taxonomy and list of observatories is available on GitHub[1]. Based on the technical descriptions and indicators outlined in this taxonomy, as well as the scope and definitions of AA and LA presented in Section 1, the relationship in Table 2 is established.

Table 2. Taxonomy of quality education assessment and alignment to AA and LA.

Taxonomy Metrics	AA/LA
1. Governance and responsibility: Implement mechanisms and standards like effective corporate governance and stakeholder engagement, and promote student well-being, facilitating a comprehensive assessment that aligns with educational objectives.	AA
2. Institutional development: Establish a clear path involving the evaluation of the mission, objectives, values, and strategic planning, which aligns with the educational model and engages the community to promote comprehensive student development.	AA
3. Quality management: Establish a culture of self-evaluation for maintenance and improvement. Implement internal quality management systems, external audits, and a commitment to maintain and enhance quality conditions driven by a dedication to self-imposed standards of excellence and facilitated by managing institutional information and performance data to enhance development and learning outcomes.	AA
4. Economic and financial management: Require a comprehensive approach that evaluates budget execution, financial performance, and resource allocation, and key indicators like sustainability, viability, and stability, emphasizing the importance of credible strategic plans, financial audits, and safeguards for student investments.	AA
5. Information management: Make information reliable and accessible, achieved through robust document and file management systems to enhance services, inform decision-making, and promote comprehensive development, learning outcomes, and informed and effective decision-making for communities and stakeholders.	AA
6. Inter-institutional networks: Create a supportive environment for institutional development through collaboration, internationalization standards, formal agreements with international institutions, and the fostering of cooperation and exchange programs.	AA
7. Physical and technological infrastructure: Recognize the impact of technological advancements on education, evaluating and managing both physical and technological infrastructure to support education services and ensuring safety and sustainability to meet institutional needs and adapt to evolving educational and technological advancements.	AA

[1] https://github.com/research-projects-info/tendon.git

Taxonomy Metrics	AA/LA
8. Non-teaching staff: Emphasize their significant contributions, necessitating adherence to standards and procedures for personnel selection, hiring, evaluation, and improvement, creating conducive working conditions and promoting staff development.	AA
9. Study plans: Focus on course design, teaching methods, learning outcomes, evaluation criteria, and research-oriented master's and doctoral programs, while primary and secondary schools emphasize curriculum diversification, high-performance standards, and alignment with the institutional curriculum.	AA, LA
10. Resources for study and learning: Include elements like information centres and virtual learning systems to be assessed through indicators, such as libraries, documentation centres, and publications, while the focus for primary and secondary schools is on managing resources that support overall development, learning achievements, and the specific needs of children and adolescents.	AA, LA
11. Management of academic processes: Consider admission, retention, and timely graduation processes and student attendance policies and support mechanisms that align with the educational model of the institution. Primary and secondary schools focus on the well-being and comprehensive development of students, with actions tailored to individual needs and pedagogical strategies aligned with the curriculum, monitoring their learning competencies and applying standards for a comprehensive learning experience.	AA, LA
12. Teachers: Focus on strengthening the teaching career through regulations and policies that enhance teaching staff skills, tenure, and promotion, aiming to establish a permanent and highly qualified teaching team. In primary and secondary schools, the emphasis is on pedagogical support, innovation, specialization, and disciplinary updates to equip teachers with the skills needed for comprehensive learning and the holistic development of students.	AA, LA
13. Health, safety, and wellbeing: For primary and secondary education, implement specialized complementary care services for children and adolescents, and for universities and technological programs, focus on safeguarding the community's well-being, enhancing student well-being, and reducing dropout rates through various support services and mechanisms.	AA, LA
14. Relationship with the surroundings: For primary and secondary schools, emphasize collaboration with families and the local community to enhance pedagogical processes and institutional identity. In universities and technological programs, focus on establishing formal mechanisms to engage with the external environment, contributing to sustainable development, and integrating environmental approaches into professional training and research.	AA, LA
15. Social inclusion: Promote equal opportunities and inclusion for historically disadvantaged groups, focusing on equal access, support for students with special learning needs among excluded groups, and the design of inclusive policies and practices.	AA, LA
16. Mediation and labour insertion mechanisms: Promote successful job placement and improve teaching-learning processes through graduate monitoring in higher education, while primary and secondary schools emphasize assessing graduation profile attainment, the satisfaction of parents and students with services, and monitoring graduate progress.	AA, LA

3.2 Alignment with the FAIR principles

We used the FAIRsFAIR Data Object Assessment Metrics (Devaraju et al., 2022). These metrics are suggested by the RDA FAIR Data Maturity Model Working Group, as well as previous efforts by the project partners, such as FAIRdat and FAIREnough, and the WDS/RDA Assessment of Data Fitness for Use checklist. These metrics were refined through expert consultation, public feedback, and practical implementation via tools like F-UJI (Devaraju & Huber, 2021). Importantly, the FAIRsFAIR Data Object Assessment Metrics adopt a partial scoring approach, meaning that even if metadata is incomplete, it is evaluated based on the extent to which each criterion is fulfilled, and partial points are awarded where applicable. For instance, to assess "Metadata includes the identifier of the data it describes," a partial score of 0.5 was assigned when "Metadata contains data content–related information (file name, size, type)" and 1 when "Metadata contains a PID or URL which indicates the location of the downloadable data content." For easier understanding of the metric, Appendix A summarizes the scoring criteria for each metric.

3.2.1 Findable and Accessible Data Sets to Data Extraction

In our evaluation of findability (F), a set of five metrics is employed: (i) the presence of a globally unique identifier for the data, (ii) the presence of a persistent identifier for the data, (iii) the inclusion of essential descriptive core elements within the metadata (including creator, title, data identifier, publisher, publication date, summary, and keywords) to enhance data discoverability, (iv) whether the metadata contains references to the data it describes, and (v) the accessibility of metadata in a machine-readable format to facilitate automated retrieval.

We considered three key factors to assess accessibility (A): (i) whether the metadata includes information regarding the access level and access conditions of the data, ensuring transparency regarding data accessibility; (ii) the availability of metadata through a standardized communication protocol, which simplifies and streamlines access to essential information; and (iii) the accessibility of the data through a standardized communication protocol, enabling efficient and standardized access to the data set itself, thus enhancing its usability and ensuring ease of access for users. The search process began with international educational bodies such as UNESCO, OECD, and the World Bank. Then, we proceeded with the Brazilian, Colombian, and Peruvian government agencies responsible for generating, maintaining, and/or sharing educational data and other initiatives in the Latin American region.

3.2.2 Interoperable and Reusable Data Sets

We focus on three metrics to evaluate interoperability (I). First, we consider the formal representation of metadata (i) by examining how comprehensive and structured the metadata is for each data set. This includes assessing whether metadata elements such as format, searchability, accessibility, time span, and location are consistently and systematically recorded. Second, we evaluate whether metadata incorporates semantic resources (ii) by capturing contextual information within the data set, such as sample details, units of measurement, and measurement tools. This semantic enrichment enhances the interoperability of the data by providing meaningful context. Additionally, we emphasize the importance of the formal representation of metadata (iii) in ensuring that metadata is structured in a standardized and machine-readable format, making it easily interpretable and exchangeable.

Our assessment of reusability (R) encompasses five metrics. Our first step is to assess whether the information is provided in a file format that is (i) both widely recognized and compatible with commonly used software, to promote accessibility and interoperability. We also consider whether data sets contain links to related entities (ii), which enhances reusability by providing connections to relevant data and information sources. Furthermore, we assess the metadata of data content (iii) to ensure that intrinsic and contextual metadata elements are documented systematically, facilitating effective data reuse. Additionally, we look for data usage licences (iv) that specify how the data can be utilized, promoting responsible and transparent data reuse. Finally, while not explicitly mentioned, we acknowledge the significance of data provenance (v) in tracking the history and source of the data, which contributes to transparency and trust, making data more reusable.

3.3 Profiling Data Observatories and Data Sets

Table 3 shows the profiling instrument that was adapted from a tool created by A. Bowers and colleagues (2022). The adjustment includes items to record information related to Section 3.1, such as description, scope, granularity of measurement, target country, and system education. Moreover, this tool includes the degree to which the variables in each data set cover the taxonomy presented in Section 3.1.3, the result of applying the FAIRsFAIR Data Object Assessment Metrics detailed in Section 3.2 and the type of data used for LA described in Section 3.3.1.

Table 3. Profiling tool for data observatories and data sets.

Feature	Feature Description
1. Observatory/data repository/data source owner	The official name of the data observatory is the agency responsible for maintaining the data repository or data source that provides the data. System that the data was accessed from.
2. Scope of initiative	International, regional (e.g., LATAM), or local (e.g., at which administrative level of the country).
3. Description of the observatory/ data repository	Provides essential information about the research facility or digital platform where the data is collected, stored, and made accessible to users. This description includes details about the observatory's purpose.
4. Data sources	Sources of data, such as government agencies, educational institutions, research organizations, or other providers.
5. File name	Original name of the file that was downloaded.
6. Dataset ID	Unique identification number for each data set.
7. Link to file online	URL from which we downloaded a version of the data set.
8. Link to file—Local computer	Hyperlink to our downloaded version of data directories when available.
9. Country coverage	Country (Peru, Brazil, Colombia) covered by the observatory/platform.
10. File format	Format of original downloaded file (pdf, csv, xls, xlsx, txt, mdb, accdb).

Feature	Feature Description
11. Granularity of measurement	Describes the levels of data granularity at which data is provided: individual student, educational institution (school, high school, institute, university), district, state, nation, region.
12. Level of system education	Primary school, secondary school, tertiary education (university undergraduate and postgraduate, technological and pedagogical training).
13. Taxonomy for Quality Education Assessment	Lists the 16 metrics and defines the level of relation the observatory or dataset has with the taxonomy: not related (0), indirectly related (1), directly related (2).
14. Indicator	According to the institution, observatory (owner of the dataset: (taxonomy1: 0, taxonomy2: (0)).
15. Data quality assessment	Only the dimensions of completeness and timeliness. While "Uniqueness" and "Accuracy" are important data quality dimensions, they are beyond the scope of this analysis because by assessing observatories and datasets from open government and well-known research and education bodies in institutions we assume that both dimensions are curated and maintained.
16. FAIR principles	FAIRsFAIR Data Object Assessment Metrics.
17. Data supports LA/AA	Type of analytics that data supports: AA, LA, AA and LA.
18. Type of data for LA	Behavioural data, discourse data, learner registry data, and biometric data. Or "not applicable" (AP) when the data only supports AA.

3.3.1 Evaluation of the Data for the Taxonomy and Identification of the Type of Data

Our focus is on recording the degree to which the variables in each observatory and data set are covered in the taxonomy. They are evaluated at three levels: (i) not related (score of 0) denotes the absence of measures for a particular indicator, (ii) indirectly related (score of 1) suggests that there are indirect measures, and (iii) directly related (score of 2) signifies comprehensive coverage of the taxonomy indicator.

An example of the data set "Higher education programs 2021—Colombia" scored 1 for the "Study plans" indicator. This data set includes a variable for the number of credits. While this measures features of the "Study plans" indicator, it does not directly measure the requirement specified by the taxonomy, which asks, among other information, about expected learning outcomes and evaluation criteria. An example of a data set that scored 2 for "Quality management" was "MEN Higher Education Institutions 2020—Colombia" because this provides variables that inform if the institution is accredited high quality and the validity of its accreditation.

To identify the types of data for LA in the data sets, we used the classification presented by Littlejohn (2022), which includes "behavioral data (how often a learner accesses a site), discourse data (what learners say or type), learner disposition data (key characteristics associated with each learner, such as how they prefer to learn) and biometrics data."

4. Results

4.1 Landscape of Observatories and Data Sets

A total of 112 data sets were identified, aligning with the 16 education quality metrics outlined in the taxonomy (Table 2). These include 10 observatories: three international (RedIndicEs, UNESCO, and OECD), two in Brazil (National Education Plan Observatory and Education Observatory of Instituto Unibanco), three in Colombia (the University Observatory in Colombia, ExE Foundation, Open Data—Ministry of Education of Colombia, and Labour Observatory for Education), and two in Peru (the Observatory of Peruvian Education and the National Observatory of Good Practices and Educational Innovation). Each country also has two statistics platforms. Additionally, there are five open data platforms: three in Brazil, one in Peru, and one in Colombia. Other data sources were derived from observatories, platforms, and official government repositories for each respective country.

93% of the data sets originate from statistics platforms, open data sources, and government website repositories, while observatories provided more concise presentations of processed indicators. Individual data sets often contained information spanning multiple education quality metrics, such as enrolment, dropout rates, graduation rates, infrastructure, and teacher statistics. For example, the National Household Sample Survey from each country's statistical platform serves as a comprehensive source for various quality metrics. The metadata analysis for 2004 to 2022 revealed that most data sets (75%) were in proprietary formats, such as XLSX, while non-proprietary formats constituted 20% in CSV and 2% in DBF formats, including SPSS and STATA.

Section 3.1.3 shows that 50% of the taxonomy metrics (e.g., “Study plans,” “Resources for study and learning,” “Management of academic processes,” “Teachers,” “Health, safety, and wellbeing,” “Relationship with the surroundings,” “Social inclusion,” and “Mediation and labour insertion mechanisms”) consider data that can support LA. However, Figure 2 shows that, while all the data sets can support AA, no data sets offered the behavioural data, learner disposition, or biometric data necessary for LA.

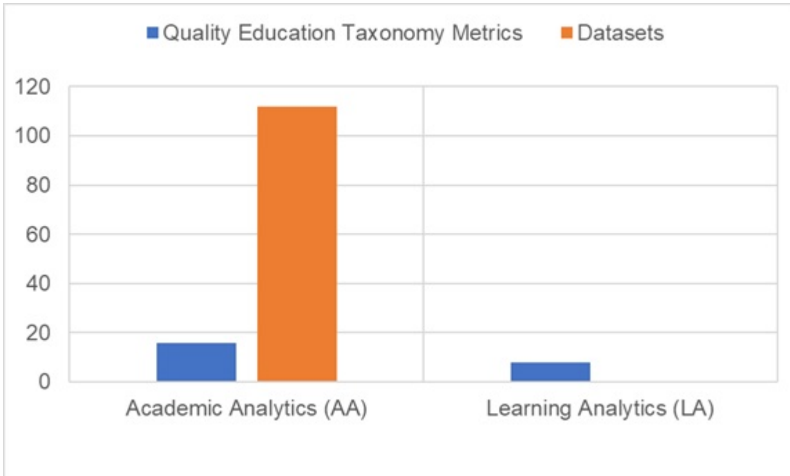

Figure 2. Comparison of quality education taxonomy metrics and data sets supporting AA and LA.

Figure 3 summarizes the data categorized by country and international observatories, covering 15 of the 16 education quality measurements outlined in the taxonomy. The “Information management” indicator was excluded due to the lack of suitable data. The blue bars denote continuous data over specific periods (e.g., “9. Study plans,” 2004–2006 and 2007–2014 for Brazil, 2004–2015 for Peru, and 2005–2010 for Colombia), and the white lines represent data available only in a particular year. For example, for the metric “13. Health, safety, and wellbeing,” Brazil has data in 2015, 2019, and 2021. While data coverage is inconsistent across many indicators, countries, and times, metrics 11 (Management of academic processes) and 12 (Teachers) offer more reliable and complete data sets across countries, organizations, and time.

Figure 4 reveals that all countries have data at the district, municipality, state/department, and country levels, but international observatories primarily collect data at the country level. While gender data is common, only Colombian data sets include ethnicity, and Brazilian data lacks age granularity. Furthermore, Figure 5 shows no Peruvian data on university postgraduate education.

Figure 6, the Sankey diagram, demonstrates how indicators (1,053 indicators in total) from international (e.g., GEO-UNESCO) and national (Colombia, Brazil, Peru) sources align with the 15 metrics of the taxonomy. GEO-UNESCO dominates with 555 indicators, heavily focused on “Management of academic processes” (440). Other prominent classes include “Teachers” (154 indicators), “Social inclusion” (129 indicators), and “Study plans” (91 indicators), reflecting key educational data availability in all countries. Country-specific data sources contribute more evenly, with Colombia (93), Peru (108), and Brazil (118) focusing on areas such as “Management of academic processes,” “Physical and technological infrastructure,” and “Mediation and labour insertion mechanisms.” “Institutional development” data is available for Brazil and Colombia, while data for “Quality management” is found only within Colombia, and “Resources for study and learning” is found within the Brazilian data set. Analysis of GEO-UNESCO and Education at a Glance (OECD) data sets reveals that data completeness for the countries studied is 66% for OECD; in comparison, UNESCO exhibits a higher completeness rate of 74% with respect to the sources and values of the indicators.

The clustered heatmap (Figure 7) illustrates the mapping of observatories, data sources, and data sets (row on the y-axis, categorized by year if the data set contains information spanning multiple years) to the taxonomy metrics (column on the x-axis, item 13 of Table 3: profiling tool for data observatories and data sets). In this graph, each data set is evaluated on a scale from 0 to 2, with varying colour intensities used for representation. White (0) signifies that no indicators were identified in the data set, light blue (1) indicates a data set containing indirectly related measures, and dark blue (2) denotes a data set that includes direct information relevant to mapping the taxonomy. Additionally, each data set was considered as a cluster, and these clusters were combined based on their similarities to create larger clusters. The clustering process involved both rows and columns, with Euclidean distance used as the distance metric and complete linkage as the method for merging clusters. Observatories and data sets show varied levels of alignment across measures, for instance, a significant portion of the data sets that cover aspects like “Mediation and labour insertion mechanisms” also include data on “physical and technological infrastructure,” “Social inclusion,” “Quality management,” and “Health, safety, and well-being.” Particularly, we could not find any data set that

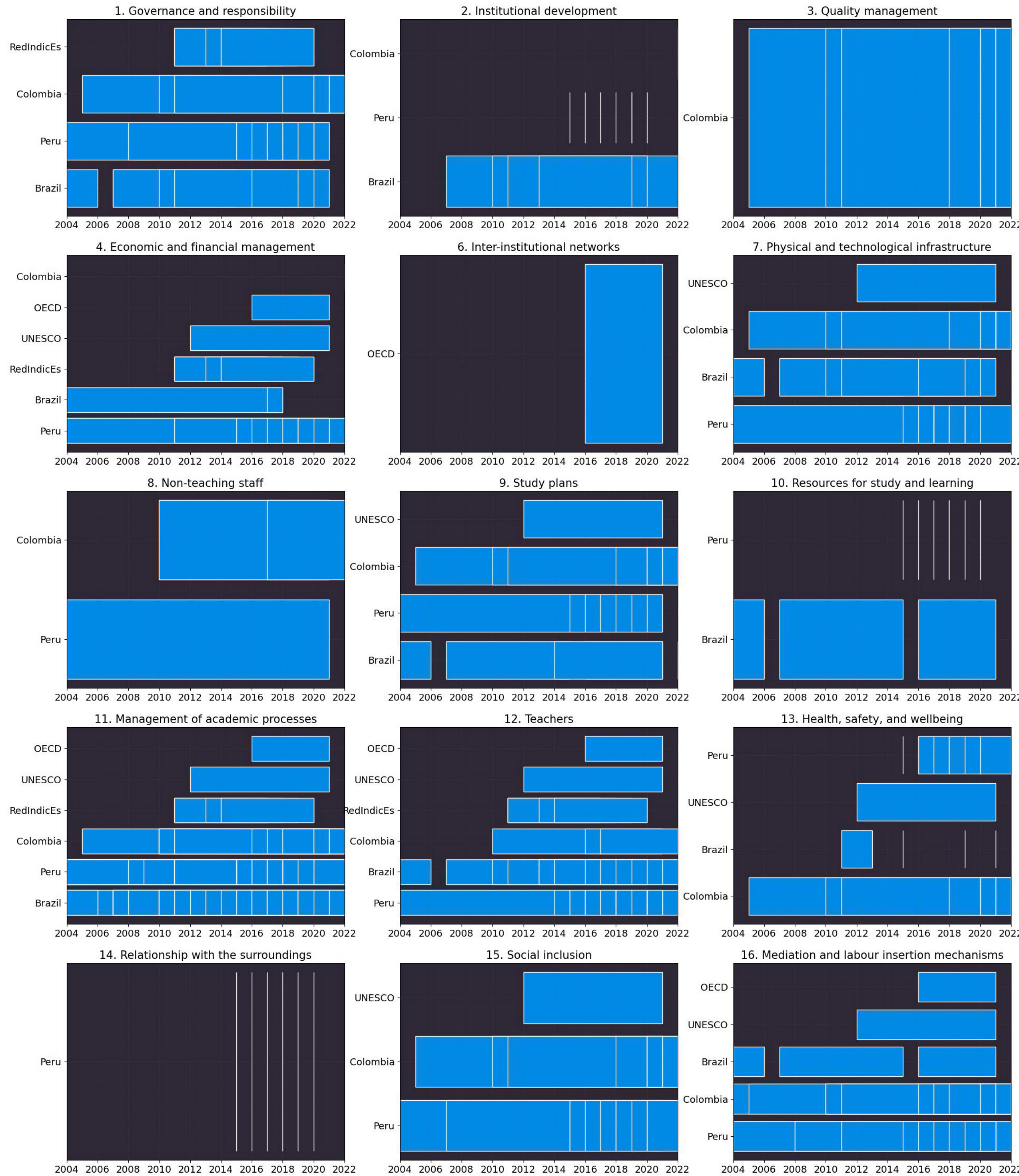

Figure 3. Availability of the data for quality education taxonomy over time (2004–2022) by observatories and countries (group observatories and data source platforms).

includes measures for "Information management." In contrast, "Management of academic processes" and "Teacher" are the most frequently covered metrics. Furthermore, "Governance and responsibility" is an aspect indirectly related to the data sets.

4.2 Alignment with FAIR Principles

Figure 8 explores our data object evaluation results. This heatmap employs the x-axis to represent FAIR measurements and the y-axis to denote the observatories and data sources, displaying compliance levels ranging from 0 (light yellow) to 100% (dark

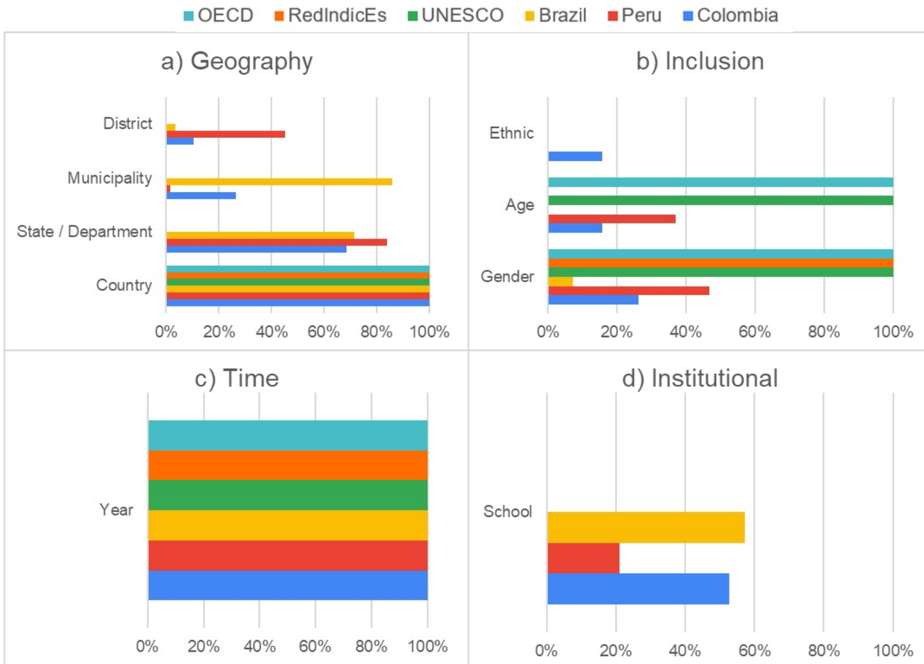

Figure 4. Level of granularity of the data by data set of the country and the international observatories (UNESCO, RedIndicEs, and OECD).

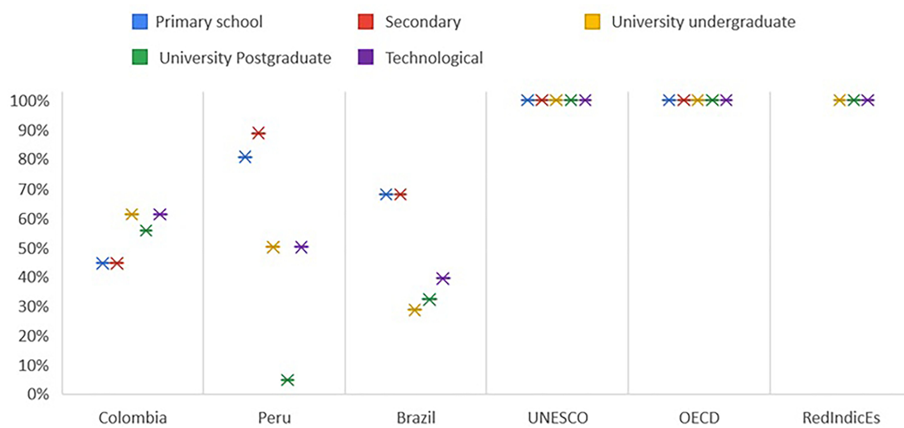

Figure 5. Scope of the level of education by observatories and country (group observatories and data source platforms) data sets.

blue) with respect to the FAIRsFAIR Data Object Assessment Metrics outlined in our research methodology. The metric F4, which assesses whether metadata is registered or indexed in a searchable resource, has achieved a high score across the board, with only two of the analyzed data sources receiving a score of 0. The low scores are explained by the sources' inability to meet the requirements related to the data identifier (IRI, URL) and metadata provision endpoint.

The findability (F) criterion generally achieved a score of 50%, while accessibility (A) reached 33% in the majority of cases. The score for (F) was impacted by a 0 rating in F3, where metadata lacked clear and explicit references to the unique identifier. This absence of a direct link between metadata files and data sets, which are typically separate, presented a significant challenge in categorizing data within our quality education measurement taxonomy, necessitating extensive exploration to locate relevant documentation. For example, UNESCO's data, which offered downloadable CSV files without comprehensive variable descriptions, or INEP and the MEN, with limited metadata on indicators but not variables, had similar issues. Peru's Open Data Education platform faced comparable challenges with unlinked and incomplete metadata.

Concerning accessibility (A), this metric evaluates the inclusiveness of data access levels. An exploration of open data repositories revealed instances where access required registration. For example, EPBM Teacher-Officials data in Colombia's

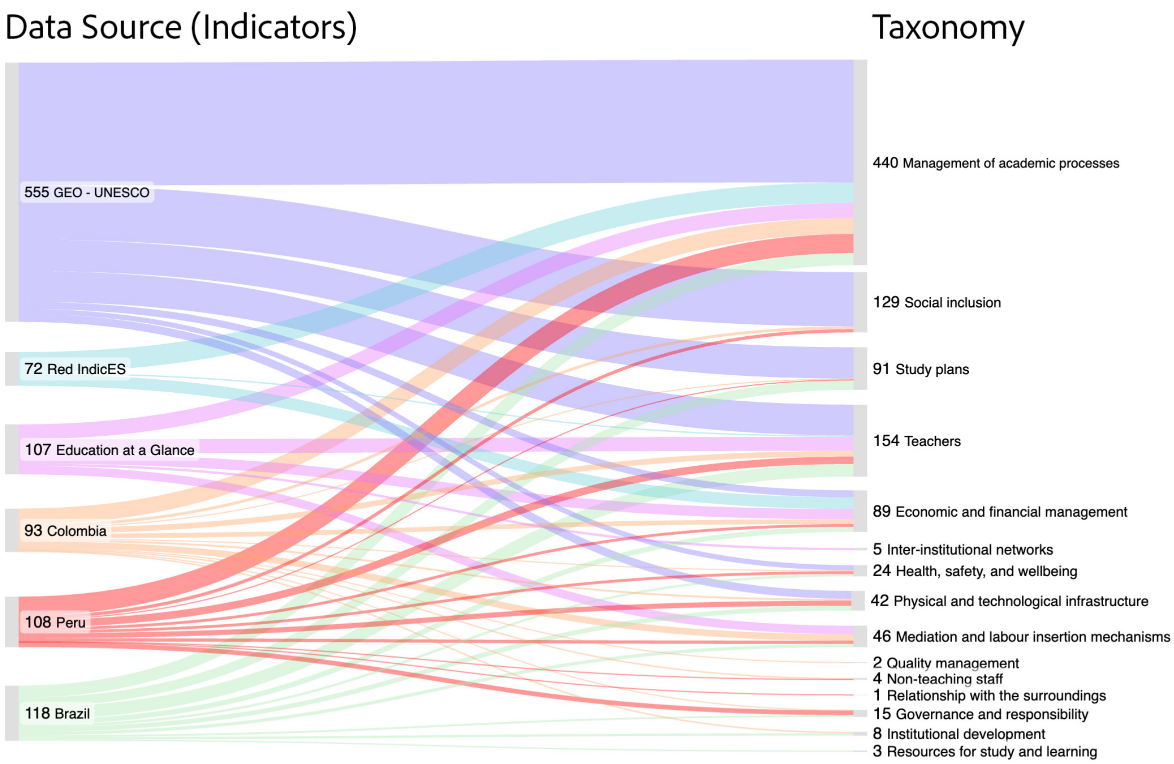

Figure 6. Sankey diagram of the proportions of indicators from international, Colombian, Brazilian, and Peruvian data sources (observatories, open data, statistics platform) mapped to the 15 classes of the Quality Education Measurement Taxonomy.

open data platform and the National Observatory of Good Practices and Educational Innovation require signing up to download information related to good practices and their systematization.

The evaluation of the interoperable (I) and reusable (R) principles, aimed at enhancing efficient data utilization and reusability, resulted in scores falling below 50%, with a few exceptional cases achieving 50% and 75%. Confirming compliance with the use of formal knowledge representation languages like RDF, RDFS, and OWL, along with their various serialization formats, which enable meaningful machine processing of domain knowledge and enhance data exchange capabilities, proved challenging for the majority of the data sources. Platforms like UNESCO's statistics offered limited metadata visibility, providing brief source information for indicators but often lacking globally unique and persistent identifiers. A similar situation was observed in Colombian, Peruvian, and Brazilian data sources, where mentions were made at the institutional level rather than specifying the data set or repository links, hindering comprehensive citation practices.

In addition, data sources consistently score poorly for R (falling below 30%). They often provide insufficient, incomplete, or unclear metadata. Not every report was accompanied by supporting documentation, and even in cases where data dictionaries were present, this documentation was frequently inadequate and devoid of details regarding the methodology used to calculate variables. This gap is a challenge for educational researchers and policymakers who seek to evaluate the relevance and applicability of the data in the specific context of quality education.

5. Discussion

The identification of 112 data sources reveals the availability of a diverse range of resources across Brazil, Colombia, and Peru, reflecting efforts to address various aspects of educational quality. The presence of international observatories like UNESCO, RedIndicEs, and OECD provides international frameworks for comparison, while national observatories and platforms offer more localized insights. However, the distribution of these resources highlights disparities in data accessibility and focus across the countries. For instance, the reliance on government repositories suggests a need for improved standardization and integration of data. The presence of five open data platforms indicates progress in enhancing public accessibility, but further efforts are required to align these data sets with international standards and ensure their utility in supporting broader educational quality analysis.

The identified data allows various educational quality metrics mainly to support AA. However, the lack of data for LA, due to its limited granularity and traceability (Lang et al., 2022), restricts the potential for advanced analyses to inform

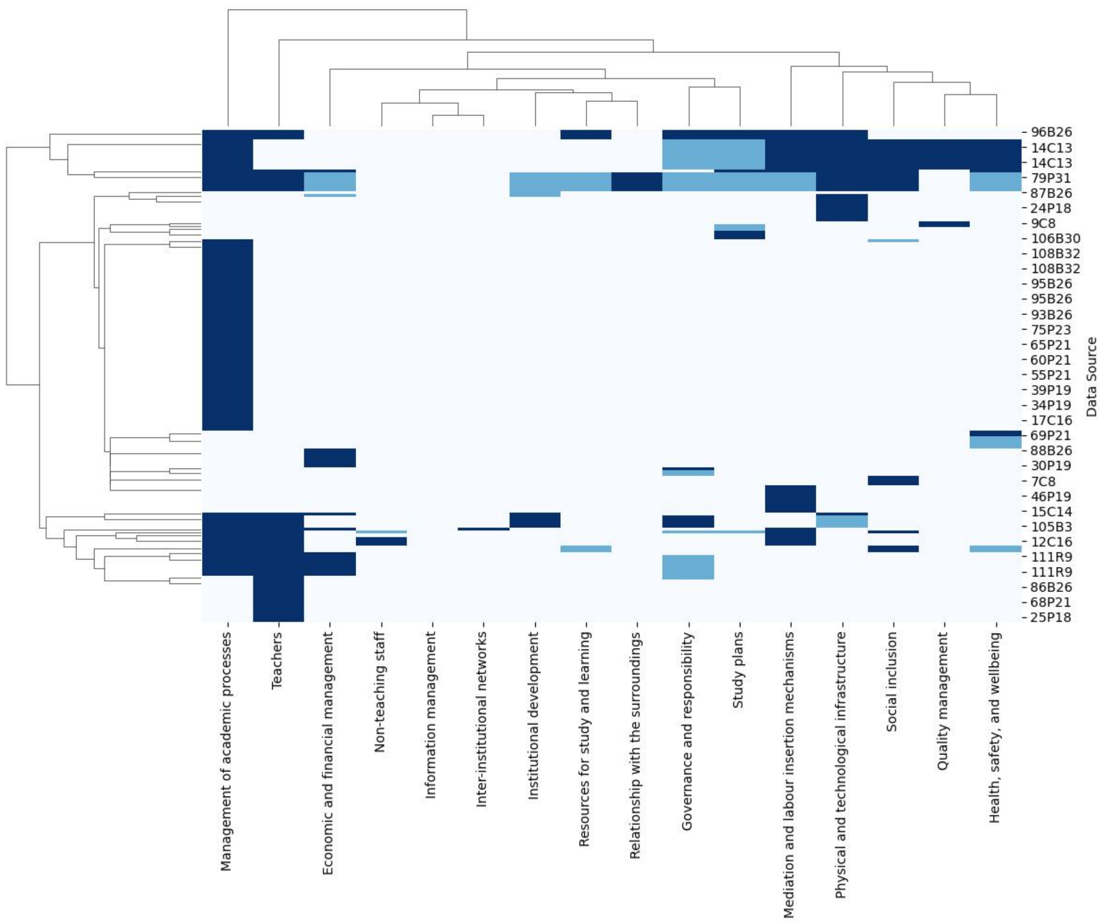

Figure 7. Clustered heatmap of observatories, data sources, and datasets, revealing alignment patterns, thematic overlaps, and gaps in coverage across 16 quality education measures (taxonomy). The x-axis displays the taxonomy metrics, while the y-axis shows data sources and data sets, which are colour-coded from white (0 = not related), to light blue (1 = indirectly related), to dark blue (2 = directly related). Clustering on both axes reveals thematic overlaps and data gaps; for instance, "Mediation and labour insertion mechanisms" often overlap with those covering "Physical and technological infrastructure," "Social inclusion," "Quality management," and "Health, safety, and wellbeing." Moreover, there is a data gap for "Information management."

educational improvements through personalized insights. This last finding aligns with previous research in LA, such as Freitas and colleagues (2024) in Brazil, Kotorov and colleagues (2024) in Colombia, and Biernacka and Huaroto (2020) in Peru; they share the limitation on the availability of data, as noted by Hilliger and colleagues (2024). To address these gaps, further investment in data collection related to learning processes and learner behaviour is essential for advancing the field of LA.

The predominant reliance on proprietary formats, the lack of suitable data for quality indicators like "Information management," and the temporal gaps and transitions in data availability (Figures 3, 4, and 5) highlight the importance of enhancing data standardization and coverage to ensure more extensive monitoring and analysis of educational quality metrics over time, country, and international level. Furthermore, the limited ethnic data, with Colombia being an exception, hinders analysis of ethnic disparities in education.

There are disparities in data coverage and quality across international and national data sets. While UNESCO and OECD data sets demonstrate relatively high completeness rates, gaps persist in addressing less commonly measured metrics, such as "Institutional development" and "Quality management," that are relevant for thorough quality education assessments. Peru's focus on highly specific indicators contrasts with Colombia's and Brazil's more expansive and interconnected measurement approaches. These variations reflect differences in national priorities and data collection strategies. Furthermore, the reliance on numerical data sets limits the exploration of rich textual data (e.g., reports, legal documents, and guidelines), which could provide complementary insights if analyzed using advanced document analysis methods. Addressing these limitations and leveraging diverse data types is essential for enhancing educational quality assessment frameworks and supporting LA goals ("What is learning analytics?", n.d.).

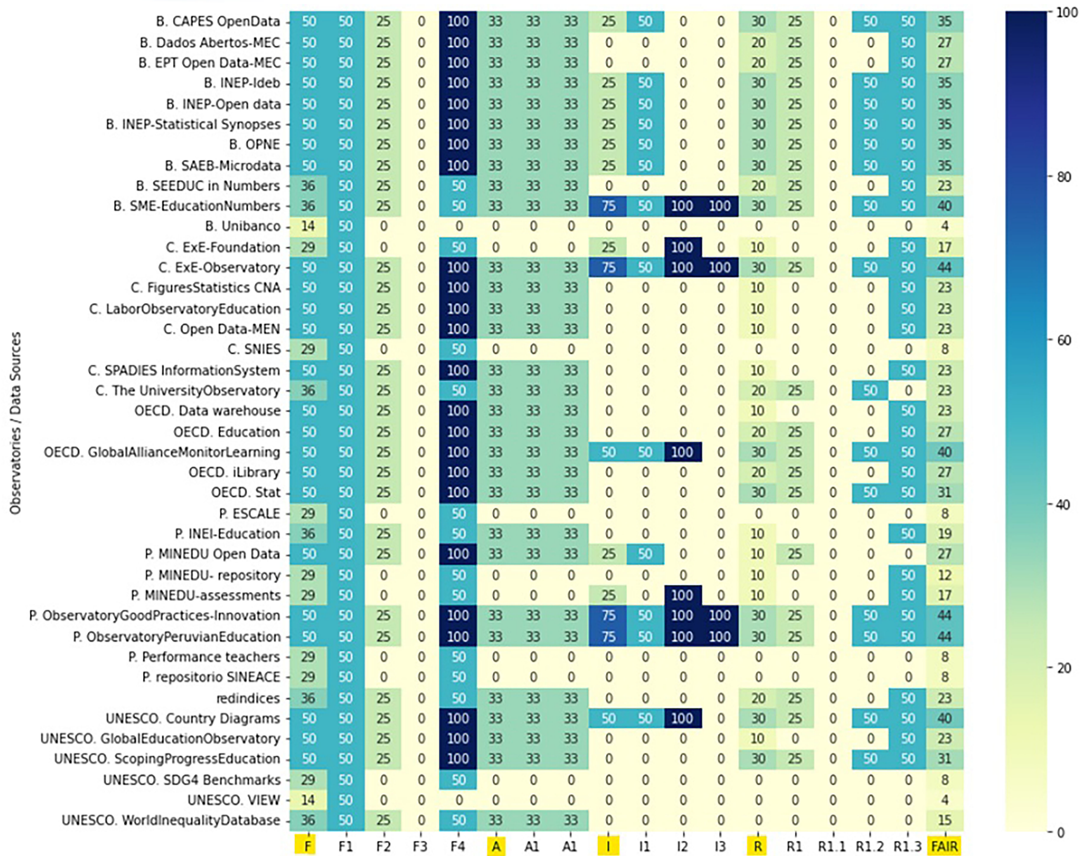

Figure 8. FAIR assessment result: Compliance heatmap (0–100%) across observatories and data sources, showing varied performance in findability (F), accessibility (A), interoperability (I), and reusability (R). The result in percentage uses the formula $score\ percent = (score\ total / score\ earned)100$. The x-axis represents individual FAIR metrics, while the y-axis lists the evaluated data sources. Colour intensity indicates compliance levels ranging from 0 (light yellow) to 100% (dark blue). R and I show most data sources scoring below 50%; F generally achieved 50% and A achieved a maximum of 33%.

Data coverage across taxonomy metrics varies (Figure 7), revealing both strengths and gaps. Metrics like "Management of academic processes" and "Teacher" are well-represented, likely due to their importance in educational quality assessments. However, despite their potential relevance to LA, data for these metrics is completely absent (Figure 2). Clustering analysis reveals thematic overlaps: data sets addressing one metric, such as "Mediation and labour insertion mechanisms," often also cover related aspects like "Physical and technological infrastructure" and "Social inclusion." While these overlaps offer opportunities for integrated analyses, they also suggest potential redundancies in data collection. Furthermore, exploring indirect coverage of metrics like "Governance and responsibility" could significantly enhance the utility of data sets for educational evaluations. Overall, these findings underscore the need for broader and more balanced data collection strategies to support a holistic understanding of educational quality.

Although we were unable to directly evaluate LA data, focusing on FAIR principles remains valuable. Data sets adhering to these principles could serve as a foundation for future AA and LA initiatives. Considering that no single institution or country has sufficient data to build large-scale AI models for learning, collaborative, networked data models are essential to pool resources and create a robust LA ecosystem (Society for Learning Analytics Research, 2024). Accessibility (A) revealed limitations, since exclusively public data sources were analyzed. Despite this, instances of restricted access were observed, such as the EPBM Teacher-Officials data in Colombia's open data platform and the National Observatory of Good Practices and Educational Innovation in Peru, which required registration for access. These findings emphasize the need for a FAIR-compliant, discipline-specific data repository with user-based access to meet LA needs (Wolff et al., 2021).

The evaluation of the I and R principles showed challenges in efficient data utilization and reusability, with scores consistently falling below 50%, except for a few exceptional cases. Ensuring compliance with formal knowledge representation languages like RDF, RDFS, and OWL, which enhance machine processing and data exchange, remains an issue for most data sources. For instance, UNESCO's statistics platform provided limited metadata visibility, offering only brief source information for indicators without globally unique or persistent identifiers. A similar issue was observed in Colombian, Peruvian, and

Brazilian data sets, where institutional-level mentions replaced data set–specific links, obstructing proper citation practices. R showed particularly poor performance, with scores below 30% across the board. This inadequacy stems from insufficient and incomplete metadata, with data dictionaries often lacking methodological details on variable computation. These problems make it hard to evaluate data relevance and applicability to quality education contexts. Addressing them requires concerted efforts to adopt global standards for metadata and improve documentation practices, which are critical for advancing LA and ensuring data usability on a broader scale.

5.1 Limitations and Future Work

This study relies mainly on publicly accessible observatories and data sets, disregarding proprietary or non-public sources that could enhance comprehension of education quality. It only considers quantitative data. Moreover, the evaluation of data adherence to FAIR principles exposes deficiencies in findability (F), accessibility (A), interoperability (I), and reusability (R), suggesting that the research scope might be restricted by data limitations due to ethical privacy regulations. Another limitation is that this study focuses on only three countries—Brazil, Colombia, and Peru—within the broader Latin American region. Although this narrows the scope, it also presents a strength in laying the groundwork for future research in other countries, enabling comparative analyses and broader regional insights.

More data is needed to support continuous observation and insight into diverse levels and educational systems in LATAM, considering the type of data required for AA and LA. This data will facilitate ongoing efforts to identify best practices, inform cross-country policy discussions, address educational disparities, and foster cross-cultural understanding. Furthermore, the implementation of FAIR principles in both public and private educational data sources is recommended as a key priority for future work applied to AA and LA. Standardizing and enhancing accessibility through these principles will not only contribute to the credibility of research findings but also promote collaborative efforts among researchers, policymakers, and educators.

Therefore, we recommend the following:

- **For policymakers:** With stakeholders' collaboration, establish and apply educational data policy and strategies aligned with the FAIR principles that integrate LA. Focus on defining actionable "rules" rather than simply recommending good practices to improve machine-actionability as recommended in FAIRsFAIR (2021):
  - Implement FAIR-aligned certification requirements for public repositories[2].
  - Ensure that research policies[3] consider the use of FAIR principles.
  - Consider the level of FAIRness and data sharing as part of a quality education assurance assessment of the educational institutions.
- **For national and international institutions:** Promote national and cross-country efforts to improve data interoperability, facilitate capacity building in data governance, and foster open data initiatives that respect privacy and ethical considerations while enabling responsible educational research in LA and evaluation of quality education:
  - Use FAIR assessment tools, for example, F-UJI[4], to audit the maturity of FAIRness in open data platforms and implement necessary improvement actions.
  - Make repositories support FAIR by developing tools, such as APIs, and share best practices and user stories.
  - Implement data validation rules to ensure consistent data formatting, accuracy, and completeness.
  - Implement services[5] supporting metadata[6] and domain-specific LA ontologies, as domain-specific requirements have to be taken into account.
  - Use guidelines and best practices for enhancing metadata quality across educational datasets.
  - Encourage closer collaboration between the institutions (private and public) responsible for educational observatories, repositories, and open platforms and the research community.
  - Foster regional collaboration on FAIR implementation challenges and emerging solutions through countries.
  - Develop technical training[7] on FAIR data principles, LA technologies[8], and data governance with certification.

[2]https://data.europa.eu/doi/10.2777/127253
[3]For example, https://www.fair-access.net.au/fair-statement
[4]https://github.com/pangaea-data-publisher/fuji
[5]"Recommendations for services in a FAIR data ecosystem": https://doi.org/10.1016/j.patter.2020.100058
[6]"A first metadata schema for learning analytics research data management": https://www.o-bib.de/bib/article/view/5735/8557
[7]"Progress toward a comprehensive teaching approach to the FAIR data principles": https://doi.org/10.1016/j.patter.2021.100324
[8]Resources: https://www.solaresearch.org/core/

- On the advance report of SDG 4[9], "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all," include recommendations on educational data availability and FAIRness to promote tools and systems for quality education improvement.

- **For researchers:** Collaborate to create a discipline-specific research data repository (Wolff et al., 2021) for large-scale LA, equipped with infrastructure and resources to support FAIR-oriented data management in the region:
  - Promote partnerships between universities, research centres, and governments[10] to harmonize data curation practices and promote shared best practices in data stewardship.
  - Contribute to the development and refinement of domain-specific metadata standards and ontologies that enhance the interoperability of educational data.
  - Develop tools (e.g., scripts, APIs) that facilitate the FAIRification and use of educational data within the research community.
  - Collaborate on contextualized guidelines[11] for FAIRness and ethical[12] uses of data in the educational field and the country of origin or the country of employment of the researchers.

6. Conclusion

We examined publicly accessible observatories and data sets on education quality in Brazil, Colombia, and Peru, focusing on quality and FAIR data principles. Our findings have important implications for developing a quality indicator system.

FAIR Principles and Data Accessibility Our assessment of the observatories and data sources was guided by the FAIR principles. The findings revealed sub-optimal results in terms of findability and accessibility, primarily due to issues related to metadata clarity and the availability of direct links to data sets. Furthermore, data sources exhibited poor scores regarding data reusability and interoperability, frequently characterized by insufficient, incomplete, or unclear metadata, highlighting the need for substantial improvements in these areas to enhance data accessibility and maximize its potential impact. FAIR data repositories are needed for AA and LA.

Landscape of Observatories and Data Sets We identified 112 data sets from observatories, statistics platforms, and open data platforms. Observatories offer more concise presentations of processed indicators, while individual data sets often encompass information spanning multiple quality metrics, making them comprehensive repositories of quality metrics. Moreover, most of the data sets were in XLSX format, followed by CSV and DBF, with some offering additional data suitable for statistical software.

Data Granularity and Coverage Data sets typically offer data at the country, year, and gender level, encompassing all education levels (except Red IndicES, which focus on higher education). Most lack ethnicity data, raising concerns about inclusivity. Additionally, Peruvian postgraduate data appears limited compared to others. While national surveys offer valuable administrative data, they, along with existing LA research, lack the granular behavioural, learner disposition, or biometric data necessary for LA applications.

Data Completeness and Focus There is significant emphasis on factors related to the management of academic processes, social inclusion, study plans, and teacher-related metrics. However, a gap in the coverage of the data is observed for other critical quality indicators, notably "Institutional development" and "Quality management," with these aspects predominantly accessible through specific country data sets, suggesting the need for more comprehensive and standardized data across these areas.

Data Quality Challenges Although a substantial volume of data is accessible, variations in data processing and focus in the three countries present challenges. Data quality concerns arise, with data completeness varying from 66% for OECD to 74% for UNESCO. These results in data quality and methodology emphasize the complexities in data interpretation, underscoring the need for standardized practices to enhance comparability and clarity in the analysis of educational data.

[9] https://sdgs.un.org/goals/goal4
[10] For instance, government institute: https://www.go-fair.org/national-offices/go-fair-brazil/
[11] https://www.go-fair.org/resources/go-fair-materials/materials-for-countries/
[12] "Technological frameworks on ethical and trustworthy learning analytics": https://doi.org/10.1111/bjet.13236

Data Gaps and Challenges Our research has brought to light the possibility of missing data sets, particularly in critical areas such as "Information management," "Resources for study and learning," and "Institutional development," which raises concerns about the accessibility of vital data. Despite our dedicated efforts, uncertainties persist regarding the existence of these data sets, emphasizing the ongoing challenge of comprehensively identifying data for research and analysis within the realm of quality education assessment. The type of data found supports AA but not LA, which underscores the limitations of observatories and open data sets in supporting LA.

In conclusion, this research highlights and advocates for the need to improve data quality and accessibility; fill the gaps in types of data (e.g., behavioural, discourse, learner disposition, and biometric data) for LA; and improve adherence to FAIR principles in the education sector of Brazil, Colombia, and Peru. Addressing data gaps and enhancing data clarity and accessibility are crucial for facilitating research and policy-making to improve the quality of education in these countries. Moreover, international organizations and individual countries may work toward standardizing data collection methodologies and improving data quality to make cross-country comparisons more meaningful.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The publication of this article received financial support from the Australian National Centre for Latin American Studies (ANCLAS) and partially from the National Council for Scientific and Technological Development (CNPq)—Brazil—Process 305436/2021-0.

References

Bautista Godínez, T., Castañeda Garza, G., Pérez Mora, R., Ceballos, H. G., Luna De La Luz, V., Moreno-Salinas, J. G., Zavala-Sierra, I. R., Santos-Solórzano, R., Moreno Arellano, C. I., & Sánchez-Mendiola, M. (2024). Perspectives and opportunities for learning analytics integration: A qualitative study in Mexican universities. *Journal of Learning Analytics*, *11*(1), 49–66. https://doi.org/10.18608/jla.2024.8125

Biernacka, K., & Huaroto, L. (2020). Learning analytics in relation to open access to research data in Peru. An interdisciplinary comparison. *Proceedings of III Conferencia Latinoamericana de Analíticas de Aprendizaje* (LALA 2020), 1–2 October 2020, Cuenca, Ecuador.

Bloom-Weltman, J., Honey, R., Meholick, S., & Fettro, M. (2018). *Profile of state data capacity in 2018: Statewide longitudinal data systems (SLDS) survey descriptive statistics* (tech. rep.). National Center for Education Statistics. https://nces.ed.gov/Pubs2021/2021126/

Bowers, A., Choi, Y., Shi, H., Sun, F., Sun, H., Williams, J., & Weeks, S. (2022). *Mapping 16 equity indicators to the New York City Schools Public Datasets: Access, opportunity, and outcomes* (tech. rep.). Columbia University. https://doi.org/10.7916/A143-AQ05

Bowers, A. J., Bang, A. H., Pan, Y., & Graves, K. E. (2019). *Education leadership data analytics (ELDA): A white paper report on the 2018 ELDA Summit* (tech. rep.). Columbia University. https://doi.org/10.7916/D8-31A0-PT97

Bowers, A. J., & Choi, Y. (2023). Building school data equity, infrastructure, and capacity through FAIR data standards: Findable, accessible, interoperable, and reusable. *Educational Researcher*, *52*(7), 0013189X231181103. https://doi.org/10.3102/0013189X231181103

CNA. (2022). CNA—Sistema Nacional de Acreditación en Colombia. Retrieved October 10, 2023, from https://www.mineducacion.gov.co/CNA/1741/article-186365.html

Congreso de la República. (n.d.). Ley N.° 28044. Retrieved October 12, 2023, from https://www.gob.pe/institucion/congreso-de-la-republica/normas-legales/118378-28044

Devaraju, A., & Huber, R. (2021). An automated solution for measuring the progress toward FAIR research data. *Patterns*, *2*(11), 100370. https://doi.org/10.1016/j.patter.2021.100370

Devaraju, A., Huber, R., Mokrane, M., Herterich, P., Cepinskas, L., de Vries, J., L'Hours, H., Davidson, J., & White, A. (2022, April). *FAIRsFAIR data object assessment metrics* (tech. rep.) (Version Number: 0.5). Zenodo. https://doi.org/10.5281/ZENODO.6461229

Eniceia, M., & Fabiana, C. (2015). National Observatory on Special Education: Network study about inclusive education in Brazil. *Open Journal of Social Sciences*, *3*(9), 60–64. https://doi.org/10.4236/jss.2015.39009

European Commission. Joint Research Centre. (2016). *Research evidence on the use of learning analytics: Implications for education policy* (tech. rep.). Publications Office of the European Union. Retrieved June 11, 2024, from https://data.europa.eu/doi/10.2791/955210

FAIRsFAIR. (2019, February). *The project*. Retrieved October 10, 2023, from https://www.fairsfair.eu/the-project

FAIRsFAIR. (2021, March). *Policy recommendations and support programme*. Retrieved April 7, 2025, from https://www.fairsfair.eu/policy-recommendations-and-support-programme

Freitas, E., Fonseca, F., Garcia, V., Pontual Falcão, T., Marques, E., Gašević, D., & Ferreira Mello, R. (2024). MMALA: Developing and evaluating a maturity model for adopting learning analytics. *Journal of Learning Analytics*, *11*(1), 67–86. https://doi.org/10.18608/jla.2024.8099

Fujikawa, C. (2015). The Brazilian educational system. *The Brazil Business*. Retrieved October 12, 2023, from https://thebrazilbusiness.com/article/the-brazilian-educational-system

Glick, D., Cohen, A., Festinger, E., Xu, D., Li, Q., & Warschauer, M. (2019). Predicting success, preventing failure. In D. Ifenthaler, D.-K. Mah, & J. Y.-K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 249–273). Springer International Publishing. https://doi.org/10.1007/978-3-319-64792-0_14

Hilliger, I., Ceballos, H. G., Maldonado-Mahauad, J., & Ferreira, R. (2024). Applications of learning analytics in Latin America. *Journal of Learning Analytics*, *11*(1), 1–5. https://doi.org/10.18608/jla.2024.8409

ICFES. (2012). Sistema Nacional de Información de Evaluación Educativa—Resultados. Retrieved October 30, 2023, from https://www2.icfesinteractivo.gov.co/result_ecaes/sniee_ind_resul_ecaes.htm

Inusah, F., Missah, Y. M., Najim, U., & Twum, F. (2022). Data mining and visualisation of basic educational resources for quality education. *International Journal of Engineering Trends and Technology*, *70*(12), 296–307. https://doi.org/10.14445/22315381/IJETT-V70I12P228

Jamil Cury, C. R. (2022). Acreditação na Educação Superior no Brasil: Notas a partir dos dispositivos legais da educação no Brasil. *Revista Historia de la Educación Latinoamericana*, *23*(37). https://doi.org/10.19053/01227238.13979

Karakhanyan, S., & Stensaker, B. (2020). *Global trends in higher education quality assurance: Challenges and opportunities in internal and external quality assurance*. Brill. https://brill.com/edcollbook/title/58971

Khan, B. H., Corbeil, J.-R., & Corbeil, M. E. (2018). *Responsible analytics and data mining in education: Global perspectives on quality, support, and decision making*. Taylor & Francis Group. https://www.routledge.com/Responsible-Analytics-and-Data-Mining-in-Education-Global-Perspectives-on-Quality-Support-and-Decision-Making/Khan-Corbeil-Corbeil/p/book/9781138305908

Khan, N., Thelwall, M., & Kousha, K. (2022). Are data repositories fettered? A survey of current practices, challenges and future technologies. *Online Information Review*, *46*(3), 483–502. https://doi.org/10.1108/OIR-04-2021-0204

Kotorov, I., Krasylnykova, Y., Pérez-Sanagustín, M., Mansilla, F., & Broisin, J. (2024). Supporting decision-making for promoting teaching and learning innovation: A multiple case study. *Journal of Learning Analytics*, *11*(1), 21–36. https://doi.org/10.18608/jla.2024.8131

Kumar Verma, K., & Shrivastava, N. (2015). Exploring role and associated challenges of data mining technique and it's [sic] implementation in e-governance. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *4*(8). Retrieved June 23, 2025, from https://www.researchgate.net/publication/319042146_Exploring_Role_and_Associated_Challenges_of_Data_Mining_Technique_and_it's_Implementation_in_E-Governance

Lang, C., Siemens, G., Wise, A. F., Gašević, D., & Merceron, A. (Eds.). (2022). *The handbook of learning analytics* (2nd ed.). SoLAR. https://doi.org/10.18608/hla22

Littlejohn, A. (2022). Professional learning analytics. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The handbook of learning analytics* (2nd ed., pp. 141–151). SoLAR. https://doi.org/10.18608/hla22.014

Lunga, D., Hänsch, R., Verma, U., Pacifici, F., Percivall, G., & Ullo, S. L. (2023). ARD, FAIR Earth observation principles, data fusion: Where are we and where do we need to go? In *Proceedings of the 2023 IEEE International Geoscience and Remote Sensing Symposium* (IGARSS 2023), 16–21 July 2023, Pasadena, California, USA (pp. 1122–1125). IEEE. https://doi.org/10.1109/IGARSS52108.2023.10281577

Mandinach, E. B., & Jimerson, J. B. (2022). Data ethics in education: A theoretical, practical, and policy issue. *Studia Paedagogica*, *26*(4), 9–26. https://doi.org/10.5817/SP2021-4-1

Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, *69*, 100842. https://doi.org/10.1016/j.stueduc.2020.100842

Mariano, A. M., Santos, M. R., Mello, T. M., & Gomes, M. M. F. (2022). University observatories as an alternative for monitoring social trends: A case study in Brazil. *Procedia Computer Science*, *214*, 1136–1143. https://doi.org/10.1016/j.procs.2022.11.288

Martínez, E., Garcia-Benadí, A., Toma, D. M., Carandell, M., Nogueras, M., & Del Río, J. (2023). An e-infrastructure for FAIR data management of underwater observatories. In *Proceedings of OCEANS 2023*, 5–8 June 2023, Limerick, Ireland (pp. 1–6). IEEE. https://doi.org/10.1109/OCEANSLimerick52467.2023.10244451

Mihaescu, M. C., & Popescu, P. S. (2021). Review on publicly available datasets for educational data mining. *WIREs Data Mining and Knowledge Discovery*, *11*(3), e1403. https://doi.org/10.1002/widm.1403

Ministério da Educação. (n.d.). Retrieved October 13, 2023, from https://www.gov.br/mec/pt-br/pagina_inicial

Ministerio de Educación Nacional. (2021). Sistema educativo colombiano. Retrieved August 28, 2023, from https://www.mineducacion.gov.co/portal/Preescolar-basica-y-media/Sistema-de-educacion-basica-y-media/233839:Sistema-educativo-colombiano

Murillo, F. J., & Cuenca, R. (2007). Construyendo consensos en torno al concepto de educación de calidad. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *5*(3). Retrieved October 10, 2023, from https://www.redalyc.org/articulo.oa?id=55130501

Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education: An integrated view. *Journal of Information Systems Education*, *31*(1), 61–71. https://aisel.aisnet.org/jise/vol31/iss1/5

OECD. (2021a, June). *Education in Brazil: An international perspective*. https://doi.org/10.1787/60a667f7-en

OECD. (2021b, June). *OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots*. https://doi.org/10.1787/589b283f-en

OECD. (2023). *Programme for International Student Assessment—PISA*. Retrieved October 30, 2023, from https://www.oecd.org/pisa/

OECD. (2024, September). *Education at a glance 2024: OECD indicators*. https://doi.org/10.1787/c00cad36-en

Proética. (2020). Auditoria ciudadana a la Transparencia Universitaria. https://www.proetica.org.pe/publicacion/informe-auditoria-ciudadana-a-la-transparencia-universitaria/

Riginos, C., Crandall, E. D., Liggins, L., Gaither, M. R., Ewing, R. B., Meyer, C., Andrews, K. R., Euclide, P. T., Titus, B. M., Therkildsen, N. O., Salces-Castellano, A., Stewart, L. C., Toonen, R. J., & Deck, J. (2020). Building a global genomics observatory: Using GEOME (the Genomic Observatories Metadatabase) to expedite and improve deposition and retrieval of genetic data and metadata for biodiversity research. *Molecular Ecology Resources*, *20*(6), 1458–1469. https://doi.org/10.1111/1755-0998.13269

Rivas, A., & Sanchez, B. (2022). Race to the classroom: The governance turn in Latin American education. The emerging era of accountability, control and prescribed curriculum. *Compare: A Journal of Comparative and International Education*, *52*(2), 250–268. https://doi.org/10.1080/03057925.2020.1756745

Sales, L., Henning, P., Veiga, V., Costa, M. M., Sayão, L. F., Da Silva Santos, L. O. B., & Pires, L. F. (2020). GO FAIR Brazil: A challenge for Brazilian data science. *Data Intelligence*, *2*(1-2), 238–245. https://doi.org/10.1162/dint_a_00046

Scavarda, A., Daú, G., Scavarda, L. F., Chhetri, P., & Jaska, P. (2023). A conceptual framework for the corporate sustainability higher education in Latin America. *International Journal of Sustainability in Higher Education*, *24*(2), 481–501. https://doi.org/10.1108/IJSHE-07-2021-0272

Shanks, R., Scharlau, B., Saevanee, H., & Stelfox, K. (2018). Human and technological resources needed to develop and sustain a city-wide educational data observatory. In B. H. Khan, J. R. Corbeil, & M. E. Corbeil (Eds.), *Responsible analytics and data mining in education* (1st ed., pp. 179–191). Routledge. https://doi.org/10.4324/9780203728703-13

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *5*, 30–32. https://doi.org/10.17471/2499-4324/195

Society for Learning Analytics Research. (2024, February). Panel 3: Where is learning analytics going?: Examining values, goals practices and perceptions ... Retrieved June 28, 2024, from https://www.youtube.com/watch?v=tbpwToTTsi8

Somogyvári, L., Bittar, M., & Hamel, T. (2023). Observatory for the history of education: Looking at the past, analysing the present and reflecting on the future—A transnational perspective. *Paedagogica Historica*, *59*(5), 837–846. https://doi.org/10.1080/00309230.2021.1962927

SUNEDU. (2015, November). El modelo de licenciamiento y su implementación en el sistema Peruano. https://repositorio.minedu.gob.pe/handle/20.500.12799/4565

SUNEDU. (2021). Modelo de renovación de licencia institucional. https://cdn.www.gob.pe/uploads/document/file/1805076/Modelo.pdf

Talamás-Carvajal, J. A., Ceballos, H. G., & Ramírez-Montoya, M.-S. (2024). Identification of complex thinking related competencies: The building blocks of reasoning for complexity. *Journal of Learning Analytics*, *11*(1), 37–48. https://doi.org/10.18608/jla.2024.8079

UNESCO. (2017). *ICCS 2016 highlights a significant increase in average student civic knowledge in 11 countries since 2009—UNESCO* (tech. rep.). UNESCO. Retrieved October 30, 2023, from https://www.unesco.org/en/articles/iccs-2016-highlights-significant-increase-average-student-civic-knowledge-11-countries-2009

UNESCO. (2021). *UNESCO recommendation on open science* (tech. rep.). UNESCO. https://doi.org/10.54677/MNMH8546

UNESCO. (2023, January). *2023 SDG4 scorecard on progress towards national SDG 4 benchmarks: Focus on early childhood* (tech. rep.). UNESCO. https://doi.org/10.54676/ZSLV3583

van Oostveen, R., Childs, E., Barber, W., DiGiuseppe, M., Percival, J., & Desjardins, C. (2019). Introducing the Global Educational Learning Observatory (GELO) and the Global Readiness Explorer (GREx): A framework and dashboard to investigate tech competence and culture. In *Proceedings of the 18th International Conference on Information Technology Based Higher Education and Training* (ITHET 2019), 26–27 September 2019, Magdeburg, Germany (pp. 1–6). IEEE. https://doi.org/10.1109/ITHET46829.2019.8937340

Varella Ehrenfried, H., Todt, E., Weingaertner, D., Erpen de Bona, L. C., Silva, F., & Didonet Del Fabro, M. (2019). Managing open data evolution through bidimensional mappings. In *Proceedings of the Sixth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies* (BDCAT 2019), 2–5 December 2019, Auckland, New Zealand (pp. 159–162). ACM. https://doi.org/10.1145/3365109.3368774

What is learning analytics? (n.d.). Retrieved June 28, 2024, from https://www.solaresearch.org/about/what-is-learning-analytics/

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wolff, I., Broneske, D., & Köppen, V. (2021). FAIR research data management for learning analytics. In A. Lingnau (Ed.), *Proceedings of DELFI Workshops 2021*, 13 September 2021, online (pp. 158–163). Hochschule Ruhr West. https://repositorium.hs-ruhrwest.de/frontdoor/deliver/index/docId/733/file/DELFI_2021_WS.pdf

World Bank. (2022). World Bank Open Data. Retrieved October 9, 2023, from https://data.worldbank.org

Zapp, M. (2021). The authority of science and the legitimacy of international organisations: OECD, UNESCO and World Bank in global education governance. *Compare: A Journal of Comparative and International Education*, *51*(7), 1022–1041. https://doi.org/10.1080/03057925.2019.1702503

1. Appendix A: Summary of the FAIRsFAIR Data Object Assessment Metrics

Table 4. FAIR assessment: metric specification and scores.

Metric	Metric Specification	Score
F. Findable		**7**
F1. (Meta)data is assigned globally unique and persistent identifiers.		2
FsF-F1-01D. Data is assigned a globally unique identifier.		
Level 1	Identifier is not resolvable but follows a UUID or HASH type syntax.	0.5
Level 3	Identifier is resolvable and follows a defined unique identifier syntax (URL, IRI).	1
FsF-F1-02D. Data is assigned a persistent identifier.		
Level 1	Identifier follows a defined persistent identifier syntax.	0.5
Level 3	Persistent identifier is resolvable (landing page can be reached).	1
F2. Data is described with rich metadata.		2
FsF-F2-01M. Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary, and keywords) to support data findability.		
Level 1	Some metadata has been made available via common web methods (embedded, typed links, content negotiation).	0.5
Level 2	Core data citation metadata is available.	1
Level 3	Core descriptive metadata is available.	2
F3. Metadata clearly and explicitly includes the identifier of the data it describes.		1
FsF-F3-01M. Metadata includes the identifier of the data it describes.		
Level 1	Metadata contains data content–related information (file name, size, type).	0.5
Level 3	Metadata contains a PID or URL which indicates the location of the downloadable data content.	1
F4. (Meta)data is registered or indexed in a searchable resource.		2
FsF-F4-01M. Metadata is offered in such a way that it can be retrieved by machines.		
Level 2	Metadata is registered in major research data registries (DataCite).	1
Level 3	Metadata is given so that major search engines can ingest it for their catalogues (JSON-LD, Dublin Core, RDFa).	2
A. Accessible		3
A1. (Meta)data is retrievable by its identifier using a standardized communication protocol.		3
FsF-A1-01M. Metadata contains access level and access conditions of the data.		
Level 1	Information about access restrictions or rights can be identified in metadata.	0.5
Level 2	Data access information is indicated by (not machine readable) standard terms.	1
Level 3	Data access information is machine readable.	1
FsF-A1-02M. Metadata is accessible through a standardized communication protocol.		
Level 3	Landing page link is based on standardized web communication protocols.	1
FsF-A1-03D. Data is accessible through a standardized communication protocol.		
Level 3	Metadata includes a resolvable link to data which is based on standardized web communication protocols.	1
I. Interoperable		4
I1. (Meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation.		2
FsF-I1-01M. Metadata is represented using a formal knowledge representation language.		
Level 2	Parsable, structured metadata (JSON-LD, RDFa) is embedded in the landing page XHTML/HTML code.	1
Level 3	Parsable graph data (RDF, JSON-LD) is accessible through content negotiation, typed links, or sparql endpoint.	2
I2. (Meta)data uses vocabularies that follow FAIR principles.		1

Metric	Metric Specification	Score
FsF-I2-01M. Metadata uses semantic resources.		
Level 1	Vocabulary namespace URIs can be identified in metadata.	
Level 3	Namespaces of known semantic resources can be identified in metadata.	1
I3. (Meta)data includes qualified references to other (meta)data.		1
FsF-I3-01M. Metadata includes links between the data and its related entities.		
Level 2	Related resources are explicitly mentioned in metadata.	1
Level 3	Related resources are indicated by machine-readable links or identifiers.	1
R. Reusable		10
R1. (Meta)data is richly described with a plurality of accurate and relevant attributes.		4
FsF-R1-01MD. Metadata specifies the content of the data.		
Level 1	Minimal information about available data content is given in metadata (resource type, links).	1
Level 2	Verifiable data descriptors (file info (size, type), measured variables, or observation types) are specified in metadata.	+1 (var) +1 (file)
Level 2.1	Verifiable data descriptors (file info (size, type), measured variables, or observation types) are specified in metadata—var.	+1
Level 2.2	Verifiable data descriptors (file info (size, type), measured variables, or observation types) are specified in metadata—file.	+1
Level 3	Data content matches measured variables or file type and size specified in metadata.	+1
R1.1. (Meta)data is released with a clear and accessible data usage licence.		2
FsF-R1.1-01M. Metadata includes licence information under which data can be reused.		
Level 1	Licence information is given in an appropriate metadata element.	1
Level 3	Recognized licence is valid, actionable, and registered at SPDX.	2
R1.2. (Meta)data is associated with detailed provenance.		2
FsF-R1.2-01M. Metadata includes provenance information about data creation or generation.		
Level 2	Metadata contains elements which hold provenance information and can be mapped to PROV.	1
Level 3	Metadata contains provenance information using formal provenance ontologies (PROV-O).	2
R1.3. (Meta)data meets domain-relevant community standards.		2
FsF-R1.3-01M. Metadata follows a standard recommended by the target research community of the data.		
Level 2	Community-specific metadata standard is listed in the re3data record of the responsible repository.	1
Level 3	Community-specific metadata standard is detected using namespaces or schemas found in provided metadata or metadata services outputs.	1
FsF-R1.3-02D. Data is available in a file format recommended by the target research community.		
Level 1	The format of the data file is open.	1
Level 2	The format of the data file is long term.	1
Level 3	The format of the data file is scientific.	1

ISSN 1929-7750 (online).