

Educational Big Data and Global Educational Equity: Opportunities, Challenges, and Strategies

Jun Jiang

Lecturer in Nantong College Of Science and Technology, China
PhD student, Varna Free University "Chernorizets Hrabar", Bulgaria

piercejiangjj@126.com

(Received: 24 March 2026, Accepted: 30 March 2026)

(7th International Conference on Contemporary Academic Research ICCAR 2026, March 25-26, 2026)

ATIF/REFERENCE: Jiang, J. (2026). Educational Big Data and Global Educational Equity: Opportunities, Challenges, and Strategies, *International Journal of Advanced Natural Sciences and Engineering Researches*, 10(3), 239-246.

Abstract – This paper explores the role of educational big data in advancing global educational equity, its associated challenges, and corresponding countermeasures. While educational big data can facilitate the precise identification of educational disparities and optimize resource allocation, it also confronts issues such as data bias, privacy risks, and technical limitations. Particularly in developing countries, data collection and application face significant obstacles. The article analyzes practical dilemmas in leveraging educational big data to promote equity, including group segmentation and precision identification, disproportionality in special education populations, and the construction of comprehensive educational databases. It proposes strategic recommendations such as integrating educational equity into top-level design, enriching data dimensions, establishing a scientific evaluation paradigm for decision-making, strengthening data governance and security safeguards, and promoting balanced global development of educational data capabilities. These suggestions aim to provide actionable references for achieving educational equity.

Keywords – Big Data; Educational Databases; Global Educational Equity

1. Introduction

In today's digital era, educational big data is emerging as a pivotal force driving educational transformation. It not only provides robust support for educational decision-making but also facilitates the realization of educational equity. However, the development of educational big data has not been without obstacles, and it faces numerous challenges in advancing educational equity. Significant gaps exist globally in the capabilities for collecting, analyzing, and applying educational data across different regions—disparities that further exacerbate educational inequities. Therefore, an in-depth exploration of the role of educational big data in promoting equity, the challenges encountered, and corresponding countermeasures holds critical importance for achieving global educational equity. This paper will elaborate in detail on the following aspects, the development and challenges of educational big databases, practical dilemmas and improvement pathways for leveraging big data to advance equity, risks and challenges in applying educational big data, and strategic recommendations for promoting educational equity. These discussions aim to provide invaluable references for realizing global educational equity.

2. Development and Challenges of Educational Big Databases

2.1. Case Studies in the Development of Educational Big Database

Globally, the construction of educational big databases has a long-established history, with the databases of the Organization for Economic Co-operation and Development (OECD) and the United States serving as representative models. These databases not only possess extensive historical depth but also feature comprehensive data collection, providing critical foundations for educational research and policy formulation.

Since its establishment in 1961, the OECD has conducted educational data collection and analysis. In its early stages, it performed international education surveys across 20 founding member states, covering areas such as education budgeting (Organisation for Economic Co-operation and Development, 2025). By the late 1990s, the scope of data collection expanded significantly. Today, its 38 member states contribute to regularly published international education reports (Taylor, Kugiya, Charran & Childs, 2023). The United States initiated educational data collection in 1867, conducted its first nationwide education survey in 1870, and established the National Center for Education Statistics (NCES) in 1962. This gradually evolved into one of the world's largest national education databases. In 1969, it launched the National Assessment of Educational Progress Project, and from the late 20th to early 21st century, systematically advanced the development of its education database. These initiatives have offered vital references for global educational big data development (National Center for Education Statistics, 2019).

2.2. Evolution of Educational Big Data and the Emergence of Equity Concerns

Educational big data has evolved from fragmented educational data to large-scale educational databases and further to integrated big data ecosystems. Early educational data primarily focused on goal attainment, performance evaluation, and outcome measurement, seldom addressing issues of educational equity. Such data, limited in volume, velocity, and variety, exhibited significant shortcomings—including crude classification, absence of sharing mechanisms, and cross-institutional collaboration barriers that hindered precise support for educational management decisions and policy innovation. Organizations and nations controlling databases failed to effectively leverage data to identify educational gaps or allocate resources to marginalized groups. Consequently, despite substantial data accumulation, improvements in education and living conditions for impoverished and vulnerable communities remained elusive. Historically, educational data even exacerbated the marginalization of racial minorities and disadvantaged groups, serving as tools to perpetuate inequality. Critical education scholar David Gillborn noted that quantitative data has often been weaponized to oppose or trivialize discussions on educational equity, becoming a "facade" that conceals problems rather than a lever for change, which reveals the instrumental contradictions within educational data (Gillborn, Warmington & Demack, 2018).

2.3. Opportunities and Challenges for Developing Countries

Educational big data holds transformative potential for developing nations yet presents formidable challenges. Educational institutions in these countries must enhance awareness, knowledge, capabilities, and technological infrastructure while increasing investment. Simultaneously, they confront deep-rooted resistance from traditional bureaucratic systems and organizational cultures. This dual challenge traps many institutions in outdated data analysis paradigms. Currently, developing countries urgently require diverse data to validate transformative initiatives aimed at enhancing educational equity and student achievement. However, shortages in human resources, funding, and technical capacity, compounded by complex administrative systems, impede their ability to emulate OECD or U.S.-style national-scale educational databases. Many countries' education systems remain at nascent stages of conceptualizing data collection, far from meeting requirements for sustained, consistent, reliable, and representative data gathering. For instance, numerous nations in South America, Africa, and Asia lack systematic mechanisms to collect and report local data beyond providing limited information to the OECD, thereby excluding them from international data-sharing platforms (e.g., the Statistical Data and Metadata eXchange). Inadequate basic network infrastructure, especially in rural areas, renders reliable and

equitable data collection nearly impossible. This severely constrains the capacity of governments, non-profits, and schools to make evidence-based decisions, obstructing educational and socioeconomic improvements. Thus, developing nations must strive to ride the big data wave while overcoming immense obstacles, which is a pursuit critical not only for realizing the vision of global educational equity but also as a core challenge demanding immediate attention from global education leaders and policymakers.

3. Practical Dilemmas and Improvement Pathways for Educational Big Data in Advancing Equity

The practice of leveraging educational big data to advance educational equity is constrained across five dimensions, human resources, financial capacity, technological capability, geographical coverage, and socio-political environments. These factors impede the acquisition of accurate data and the realization of equity. Resolving these issues is essential for cultivating awareness of educational equity, formulating scientific collection strategies, and developing precise survey instruments. The effectiveness of database tools depends on the robustness of data collection and the breadth of variables, and enhancing efficacy requires expanding survey instruments to collect detailed characteristic variables of specific populations. Even in countries with mature data collection systems like the United States, survey tools often fail to adequately identify particular groups. Nevertheless, precise collection of identity data remains crucial for promoting educational equity. Drawing from empirical lessons, multifaceted efforts must be undertaken to advance this work.

3.1. Necessity of Group Segmentation and Precision Identification

Gillborn et al. point out that excluding marginalized groups from data collection and analysis exacerbates their marginalization, subjecting them to the predicament of "surveillance without services." (Gillborn et al., 2018). Overly broad racial or ethnic classifications obscure substantial cultural and socioeconomic disparities within groups, hindering accurate understanding of equity gaps. For instance, the United States frequently categorizes Asian Americans as a monolithic group, neglecting significant internal achievement variations. Such classification reinforces the false myth of the "model minority," distorting perceptions of reality. Similar issues exist in gender data collection. Although women now outnumber men in U.S. higher education, males in at least 18 countries globally exhibit lower basic education completion rates than females, highlighting the complexity of educational inequality. In the U.S., Latino males have the lowest high school graduation rates, while Latina females surpass 60% in higher education completion. In response, initiatives like Project MALES provide mentorship to youth of color, enhancing their educational opportunities and graduation rates (Stewart, 2022). Therefore, newly established databases must design more granular survey tools to sufficiently segment race, ethnicity, ethnic groups, and gender identities. This will authentically reflect intra-group diversity, lay the foundation for equitable resource allocation and precise gap identification and resolution, and prove critical for achieving global educational equity.

3.2. Disproportionality in Special Education Populations

Since the 1970s, disproportionality in special education populations has gained prominence. U.S. studies indicate that minority students with disabilities are overrepresented in special education systems (NEA Education Policy and Practice Department, 2008). Historically, systemic discrimination against students from multicultural backgrounds, linguistic minorities, and those with disabilities has been the primary cause. Although the U.S. amended the Individuals with Disabilities Education Act (IDEA) to address this, nationwide systemic data support remains insufficient. Mechanisms for collecting statistics on race, ethnicity, and disability status require refinement. Research by Alfredo Artiles and Stanley Trent reveals that existing studies inadequately examine historical and socio-structural factors in special education issues and these factors are essential for formulating policies targeting marginalized groups (Artilesaj, 1994). Van Roekel calls for collaborative action among policymakers and stakeholders to reduce disproportionality through large-scale national data collection (Goldin, 1999). Currently, disproportionality research predominantly focuses on the U.S., with insufficient international attention.

Other countries should prioritize such research to support the development of equitable education policies for vulnerable disabled populations.

3.3. Integrating Key Variables for Comprehensive Educational Databases

To achieve equitable educational outcomes, researchers must integrate key variables, such as income levels, educational attainment, and family structure, into comprehensive educational databases, in addition to collecting basic data on race, ethnicity, gender, and disability status. Household income serves as a core indicator of socioeconomic position. Longitudinal tracking studies demonstrate that income data can expose inequalities in educational resource allocation, including disparities in teacher quality, facilities, transportation, and nutrition. These factors directly impact students' educational experiences and academic achievements. Collecting stratified income data enables policymakers to precisely identify service gaps and optimize resource distribution. Although correlated with income, educational attainment possesses independent data value that cannot be overlooked. Enrollment and graduation rates from junior high to higher education constitute key metrics for assessing equity in educational opportunities. Integrating multidimensional data on race, ethnicity, and parental educational backgrounds provides precise evidence for narrowing intergenerational education gaps. Family structure exerts profound influence on educational opportunities. Family stability is closely linked to children's educational outcomes. Students in non-traditional family structures (e.g., foster care, adoption, or divorced parents) face disadvantages due to lack of stable nurturing support. Establishing databases linking family structure to educational performance aids in identifying systemic barriers for vulnerable groups.

4. Risks and Challenges in Applying Educational Big Data

Although educational big data can drive educational innovation, its application entails significant technical limitations and ethical dilemmas. If appropriate data types and structures are not selected according to regional needs, and the following issues remain unaddressed, not only will educational equity goals prove elusive, but institutions may also descend into data governance quagmires. Given that big data projects typically require long-term investment, builders, users, and policymakers must proactively recognize these challenges.

4.1. Data Overload as a Governance Challenge

Against the backdrop of rapidly evolving digital education, data overload is emerging as a novel challenge for educational governance. While massive datasets offer potential support for educational decision-making, the "three Vs" of big data, Volume, Velocity, and Variety, impose substantial pressure on educational organizations, particularly resource-constrained institutions in developing countries. This pressure manifests in three primary dimensions.

Firstly, a severe imbalance exists between data analysis capabilities and data scale. Due to shortages of specialized technical personnel, many educational institutions face "data dormancy". Although, they persistently collect vast educational data yet fail to conduct effective analysis or value extraction.

Secondly, critical data management standards are lacking. Resource-deprived schools universally grapple with blind data collection, uncertain about which core indicators to capture, unable to determine appropriate collection scales or optimal analysis timelines, and consequently trapped in a vicious cycle of "collecting for collection's sake."

Thirdly, data infrastructure and sharing mechanisms remain underdeveloped. Educational organizations lack sufficient physical or cloud storage space and have failed to establish unified data encoding and visualization systems. Consequently, stakeholders struggle to accurately interpret and effectively utilize data. These systemic deficiencies ultimately plunge educational organizations into "analysis paralysis" within the data deluge, unable to halt futile data accumulation yet incapable of initiating meaningful value mining.

4.2. Dilemmas of Data Security and Privacy Protection

In educational big data applications, data security and privacy protection constitute core challenges for educational organizations, especially in regions enforcing stringent data protection regulations (e.g., jurisdictions under the EU's General Data Protection Regulation framework). Educational institutions must ensure lawful data collection and implement robust informed consent mechanisms to safeguard the confidentiality and compliance of educational data, particularly sensitive student information. However, technological advancement far outpaces regulatory updates, and the absence of unified standards governing the storage, processing, and access of learning data perpetuates risks of data misuse. Regulatory lag subjects educational institutions to dual pressures of privacy protection and cybersecurity when advancing data-driven decisions.

Student privacy protection is especially acute. Most learning behavior data constitutes protected personal information requiring specific legal justifications for collection. Yet precision learning analytics necessitates continuous student data harvesting, while students generally lack willingness to provide such data. This reflects the complexities of data permission management and the implementation challenges of informed consent mechanisms, potentially triggering trust crises among educational stakeholders.

Furthermore, data policy transparency poses a deeper challenge. Many members of educational organizations lack awareness and interest in big data, fostering a "black-box effect" in data collection. This reduces stakeholder buy-in toward big data policies, cultivates a psychological perception of "being surveilled", and ultimately diminishes teaching efficacy, learning engagement, and organizational belonging.

4.3. Dehumanizing Tendencies in Big Data Application

In educational practice, although big data cannot function without human involvement, it risks being dehumanizingly applied, precipitating multiple issues. Overreliance on big data for decision-making strips educational participants of autonomy and efficacy. This may lead to interpersonal alienation, reducing face-to-face interactions and in-depth discussions, thereby inducing social isolation or psychological distress among students. Individual value may be diminished as educational participants are reduced to "data points" rather than vibrant individuals, eroding teacher-student trust.

Homogenization risks follow that data-driven decisions may overlook individual differences, treating students as "typical" cases subjected to uniform approaches. Educational relationships may become alienated—exemplified by the "datafication of learners" phenomenon in Australia's education system—weakening organic connections between students and teachers, schools and communities, and marginalizing certain groups (Buchanan & Mcpherson, 2019). Thus, the application of big data in education must vigilantly guard against dehumanizing tendencies, preventing technological rationality from overshadowing educational humanistic care.

4.4. Data-Driven Decision-Making as a Barrier to Equity

Data-driven educational decisions harbor limitations that may obstruct educational equity. The incompleteness and authenticity of data constitute the primary concern: databases struggle to fully and accurately reflect each learner's reality. Decisions based on incomplete or biased data yield non-targeted interventions. For example, tacit knowledge exchanged orally between teachers and students resists standardized capture and quantification. Moreover, data collection tools and methods, influenced by existing power structures and values, may reinforce inequalities.

Secondly, contextual applicability poses a prominent issue. The effectiveness of big data projects hinges on alignment with local educational needs, cultural backgrounds, and institutional environments. In transnational projects, applying a "one-size-fits-all" data approach not only fails to solve local problems but also wastes resources and exacerbates educational disparities.

Furthermore, the risk of data overgeneralization cannot be ignored. When data models from specific educational contexts are broadly applied, they may mask the differentiated needs of diverse groups, causing policymakers to neglect non-quantifiable yet vital educational factors. This reduces educational equity to a technical facade. Therefore, in advancing educational digitalization, it is imperative to

recognize data as merely a tool—not an end. Genuine educational equity must transcend data surfaces to address the ineffable essence of education and diverse needs.

4.5. Impediments to Educational Organizational Change

As bureaucratic institutions, educational organizations often cling to tradition, resisting change and innovation, thereby fostering conservative organizational cultures. This manifests in three key aspects, insufficient human and financial investment, lack of strategic big data planning, and conservative leadership culture. Research indicates that the global education sector lags behind other industries in data application. A core challenge for educational institutions is competing with the private sector for quantitative analysis and machine learning talent—critical for big data management and analysis (Macfadyen, Dawson, Pardo & Gasevic, 2014). Achieving educational equity requires data continuity, but until computerization and automation of data collection and analysis are realized, this process remains dependent on stable, high-caliber teams. Such teams must both discern the complex causes of educational inequality and design targeted solutions—underscoring the inherently human-resource-intensive nature of data work.

However, high staff turnover plagues the education sector: frequent mobility among K-12 teachers, university faculty, and administrators, compounded by multi-tiered accountability demands, disrupts data operations. This personnel flux introduces the "human variable" into complex big data systems, impeding progress toward educational equity.

5. Strategic Recommendations for Leveraging Educational Big Data to Advance Equity

Through the preceding analysis, it becomes evident that leveraging educational big data to promote educational equity is neither naturally occurring nor rapidly achievable. Therefore, the following strategies must be adopted in constructing and applying educational big data to harness its advantages while mitigating potential risks.

5.1. Integrate Educational Equity into Top-Level Design

When building and utilizing educational big databases, promoting educational equity should be incorporated as a core objective within top-level design and strategic planning. Currently, some educational big databases fail to effectively serve educational equity, thus, when establishing new databases or retrofitting existing ones, focused attention must be given to equity-related issues and associated reform practices. Internationally, the European Union addressed historical fragmentation of national education data by establishing a data-sharing mechanism through its "Education and Training 2020" strategy (European Commission, 2013). India enhanced its educational data collection system, narrowing urban-rural education gaps and advancing equity (National Informatics Centre, 2008). Educational institutions need to establish long-term development mechanisms, formulate medium-to-long-term plans, secure resource investments, assemble professional teams, and standardize workflows and protocols. This ensures sustainable development of educational big databases and maximizes their role in promoting educational equity.

5.2. Enrich Data Dimensions and Collection Metrics

Constructing educational big databases requires expanding multidimensional variables—including race or ethnicity, gender, family background, and learning styles—and establishing a multilevel indicator system encompassing student characteristics, learning processes, and teaching environments. Simultaneously, big data technologies should be leveraged to collect granular behavioral data such as learning interactions, participation patterns, and biometric indicators. Through multidimensional data fusion, comprehensive evaluations of teaching efficacy differentials and ed-tech tool applicability can be conducted. This serves personalized learning path design, equity issue diagnosis, and teaching assessment optimization, thereby advancing scientific and precise educational decision-making. Ultimately, it fosters continuous

improvement in education quality, facilitates equity realization, and lays a solid data foundation for high-quality educational development.

5.3. Establish a Scientific Evaluation Paradigm for Educational Decision-Making

Educational decision-making necessitates a scientific evaluation paradigm to avoid reliance on partial data or power-based preferences. Big data requires deep analysis, contextual interpretation, and policy translation. Data-sharing agreements must be established to promote collaborative integration and cross-system comparisons. Researchers need to integrate diverse data sources to build large-scale longitudinal education databases for systematic policy impact assessment. For example, economists Robert Barro and Jong-Wha Lee integrated UNESCO data to construct a 60-year educational database covering 146 countries, revealing that each additional year of education increases individual returns by 5–12% (Barro & Lee, 2013). California's 2021 legislation establishing the California Cradle-to-Career Data System integrates education, social service, and employment data to support tracking vulnerable populations and optimizing policies, demonstrating cross-sectoral collaboration potential (WestEd, 2025).

5.4. Strengthen Data Governance and Security Safeguards

Educational organizations must collect data lawfully and compliantly, building trust with teachers, students, and stakeholders. Data governance should be refined by clarifying usage boundaries and establishing transparent management policies. Data providers should be informed about information usage and protection measures to enhance cooperation, ensuring data accuracy and validity for evidence-based decisions. Educational organizations must elevate data security capabilities by constructing integrated safeguards encompassing technology, institutional protocols, and emergency response systems. Governmental bodies should formulate security standards and provide technical support, while educational institutions reinforce cybersecurity infrastructure and conduct safety drills. Preventing data breaches maintains public trust and addresses digital security challenges.

5.5. Promote Balanced Global Development of Educational Data Capabilities

Advancing global educational equity requires prioritizing support for big data development in developing countries. These nations should improve census mechanisms, optimize collection workflows, and establish sustainable national education databases to inform policy formulation and resource allocation, thereby effectively advancing equity. Developed countries must assist their integration into international education data systems and participation in global collaboration through technical support and resource sharing. Only by achieving balanced development of data capabilities can global educational transformation genuinely benefit all nations and populations. While pursuing their own development, developed countries should shoulder obligations to promote global equity in educational data.

6. Conclusion

Under the dual challenges of global competition and socioeconomic transformation, educational organizations face unprecedented pressures. They must simultaneously adapt to profound shifts in national political-economic landscapes and respond to the urgent demands for educational equity from disadvantaged groups and marginalized communities. In this era where data is reshaping educational ecosystems, big data has transcended its role as a mere technological tool to become a strategic resource driving educational transformation. Educational organizations must urgently establish data-driven decision-making mechanisms. Through the deep mining and intelligent application of educational big data, they can achieve optimized resource allocation and enhanced teaching efficacy. This precision-oriented educational model, underpinned by robust data, will empower learners from diverse backgrounds to overcome resource constraints, seize development opportunities in global competition, and ultimately catalyze the formation of a multidimensional, multi-tiered framework for educational equity.

REFERENCES

- [1] Organisation for Economic Co-operation and Development. (2025). Our History. [online] Available at: <https://www.oecd.org/en/about/history.html>.
- [2] Taylor, ZW., Kugiya, J., Charran, C., Childs, J (2023). Building Equitable Education Datasets for Developing Nations: Equity-minded Data Collection and Disaggregation to Improve Schools, Districts, and Communities. - *Educational Sciences*, 13(4), pp. 348-351.
- [3] National Center for Education Statistics. (2019). An Overview of NAEP. Washington, DC: National Center for Education Statistics.
- [4] Gillborn, D., Warmington, P., Demack, S (2018). QuantCrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics. - *Race Ethnicity and Education*, 21(2), pp. 158-179.
- [5] Stewart, M. (2022). The Enrollment Crisis for Men Continues to worsen. - *Insight into Diversity*, [online] Available at: <https://www.insightintodiversity.com/the-enrollment-crisis-for-men-continues-to-worsen/>.
- [6] NEA Education Policy and Practice Department. (2008). Disproportionality: Inappropriate Identification of Culturally and Linguistically Diverse Children. Washington, DC: National Education Association.
- [7] Artilesaj, T (1994). Overrepresentation of Minority Student in Special Education: A Continuing Debate. - *The Journal of Special Education*, 27(4), pp. 410-437.
- [8] Goldin, H (1999). NBER Working Paper Series on Historical Factors in Long Run Growth: a Brief History of Education in the United States. Cambridge: National Bureau of Economic Research.
- [9] Buchanan, R., Mcpherson, A (2019). Teachers and Learners in A Time of Big Data. - *Journal of Philosophy in Schools*, 6(1), pp. 26-43.
- [10] Macfadyen, LP., Dawson, S., Pardo, A., Gasevic, D (2014). Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge. - *Research & Practice in Assessment*, (9), pp. 17-28.
- [11] European Commission. (2013). European Education Area: Quality Education and Training for All. [online] Available at: <https://leducation.ec.europa.eu/>.
- [12] National Informatics Centre. (2008). Statistics of School Education. New Delhi: Government of India.
- [13] Barror, J., Lee, JW (2013). A New Data Set of Educational Attainment in the World, 1950-2010. - *Journal of Development Economics*, (104), pp. 184-198.
- [14] WestEd. (2025). California Cradle-to-career Data System. [online] Available at: <https://cadatasystem.wested.org/>.