

Article

Building Equitable Education Datasets for Developing Nations: Equity-Minded Data Collection and Disaggregation to Improve Schools, Districts, and Communities

Z. W. Taylor ^{1,*} , Jase Kugiya ², Chelseaia Charran ³ and Joshua Childs ²¹ Education and Human Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA² Department of Educational Leadership and Policy, The University of Texas at Austin, Austin, TX 78712, USA³ Département de Psychoéducation, Université du Québec à Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada

* Correspondence: z.w.taylor@usm.edu

Abstract: Many studies of education engage with large datasets to attempt to solve educational problems. However, no studies have provided a systematic overview of how large datasets could be compiled with an eye toward solving educational problems related to equity, especially as it relates to racial, gender, and socioeconomic equity. This study provides a synthesis of literature and recommendations for how developing nations can learn from peers and collect, disaggregate, and analyze data in ways that promote equity, thus improving schools, school districts, and communities.

Keywords: education; datasets; big data; decision-making; developing nations; equity

1. Introduction

Over the past three decades, attributable to the rise of the Internet and digital technologies, the era of big data (data that is large and potentially hard to manage due to volume and organizational issues) has grasped the global education community [1–3]. More so now than ever, educational leaders are using large datasets to make decisions at the student, teacher, school, district, local, national, and international levels [3–6]. With access to information, there has been exponential growth in scholarship since 2000 related to data-driven decision-making and evidence-based practices, which are implemented for the purpose of improving schools and educational outcomes for students and other stakeholders [1,2,5]. In addition, this movement toward big data has been broadly global in scope.

1.1. The Role of the Organization for Economic Cooperation and Development (OECD)

In 1961, the Organization for Economic Cooperation and Development (OECD) was founded to help spur economic growth and world trade, including gathering educational data to inform how countries can guide the development of schools and school systems [7]. During its early years, the OECD administered international educational surveys to 20 founding members about topics such as educational budgeting, teacher recruitment, student engagement, and the establishment of new schools [6]. However, in the late 1990s and the early 2000s during the dot-com boom, the OECD began gathering much more comprehensive data from member nations and their schools, including the total number of educational personnel (teachers, administrators, etc.) in 1998, student count and age of enrolled students in 2000, and graduation rates and student–teacher ratios in 2005 [8]. Now, the OECD has 38 full members across the world and publishes some of the most comprehensive international education reports available [6]. The same approach to data collection and aggregation has been adopted at the continent and country levels as well.



Citation: Taylor, Z.W.; Kugiya, J.; Charran, C.; Childs, J. Building Equitable Education Datasets for Developing Nations: Equity-Minded Data Collection and Disaggregation to Improve Schools, Districts, and Communities. *Educ. Sci.* **2023**, *13*, 348. <https://doi.org/10.3390/educsci13040348>

Academic Editor: Meihua Liu

Received: 21 February 2023

Revised: 16 March 2023

Accepted: 24 March 2023

Published: 28 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1.2. The European Union (EU)

When the European Union (EU) was established in 1993, individual EU member states were tasked with administering and measuring their own educational systems, leading to several decades of misalignment between different member states and no common dataset for EU leaders to make data-driven educational decisions [3]. However, EU leaders encouraged member states to participate in surveys administered by the OECD, while creating the “Open Methods of Coordination” (OMC) for policies in social fields, including education and training” [3] [p. 989]. Later, the EU devised the “Education and Training 2020 strategy (ET2020), as part of the Europe 2020”, which prioritized cross-country collaboration and information sharing to improve educational outcomes [3] [p. 990]. This led to the creation of the European Commission, which houses educational information and data across all EU member states [9]. These organizations ultimately allowed independent researchers to explore relationships between education outcomes and its citizenry, informing how the EU could improve its interconnected educational systems [4].

1.3. India’s Ministry of Education

Similarly, in 2002, India’s Ministry of Education developed the capacity to gather data related to organizational budgets and school expenditures to better distribute resources in India’s developing areas outside of their major cities such as Mumbai, Delhi, and Bangalore [10]. Before 2002, India’s government had only released broad data to the World Bank related to government spending on education as a percentage of India’s overall gross domestic product (GDP) without any student- or school-level data [11]. As technology proliferated in India and the Indian government established more policies to emphasize data-driven decision-making, more data was able to be collected related to school enrollment growth, the establishment of new schools, and gender equity, culminating in some of India’s most comprehensive education reports in the mid-2000s [12]. When the COVID-19 pandemic rocked the world of education, India’s school closures were among the longest in the world, averaging 73 weeks per school compared to the global average of 35 weeks [13]. Contributing to this length was the fact that many of India’s public schools are entrenched in densely populated urban areas or remote rural areas with inadequate access to medical care [13]. Yet, because of India’s increasingly centralized educational data system, India was able to swiftly compile a comprehensive report targeting equity gaps among India’s most under-resourced rural schools, allowing India’s government to provide interventions and assistance, as well as guidance on how to formulate future year budgets to fill these gaps [14].

1.4. The United States

Among developed nations, the United States (U.S.) likely has the longest-standing and most comprehensive educational data collection methods and reporting structures in the world. In the United States context, governmental policies after the first Morrill Act of 1862 greatly expanded educational opportunities for the U.S. people, and the Department of Education Act of 1867 created the U.S. Office of Education, which later became the U.S. Department of Education [15]. The aim of the office was to organize educational functions at the federal level and provide resources for states to measure the educational progress of students in their schools. In 1867, the Office of Education began making early attempts at building large datasets to measure educational goals and outcomes, with the first national-level education surveys administered and data collected being largely from public grade schools in 1870 [15]. However, scholars have long lamented that better, more robust data was not collected earlier in the history of postsecondary data collection in the United States [16].

Partially owing to the success of the 1870s surveys, the second Morrill Act of 1890 greatly expanded on the federal government’s data collection program, thrusting the U.S. into the 1900s when multiple data collection and analysis efforts built upon the Second Morrill Act: the statistical program of 1920, the Vocational Rehabilitation Act of 1943, the

Information and Education Exchange Act of 1948, and the establishment of the National Center for Education Statistics (NCES) in 1962. Today, the NCES includes secondary and postsecondary data at the school, district, state, and regional levels and is one of the most robust national educational datasets in the world [15]. Moreover, the Civil Rights Era of the 1960s and President Lyndon Baines Johnson's aggressive education agenda produced many landmark education developments in the U.S., including the signing of the Elementary and Secondary School Act (ESSA) and the Higher Education Act (HEA) in 1965 [15], both of which required data reporting by schools to the federal government. These acts paved the way for the Office of Education (now known as the Department of Education) to begin administering the National Assessment of Educational Progress in 1969, the "largest nationally representative and continuing assessment of what students in public and private schools in the United States know and are able to do in various subjects" [17] (para. 1). To date, it remains the largest and most comprehensive collection and report of big education data in the United States and the world.

Decades later, the United States developed even more formal attempts to compile large education datasets, introduced in 1990 with the advent of the National Education Goals Panel pursuant to a Congressional mandate under President George H. W. Bush [18]. The aim of the panel was to annually report on national and state educational progress toward the National Education Goals adopted by the President and the nation's governors, as well as requirements by the U.S. Office of Management and Budget for data documenting the effectiveness of federal programs both in and outside of education under the Government Performance and Results Act (GPRA) of 1994 [18]. More recently, the American Recovery and Reinvestment Act (2009) indicated that federal education officials sought to ensure that data and evidence are used to inform policy and practice [19]. The Act provided USD 10 B to "help local educational agencies hire, retain, or rehire employees who provided school-level educational and related services", including bolstering data collection and analysis initiatives related to the profession of education in the United States [19] (para. 1).

2. Issues with Big Data

Many developed nations (developed nations defined as sovereign states with a high quality of life and high Human Development Index per the International Monetary Fund) gather high-quality data to make informed educational decisions [20,21]. However, many developing nations do not have the resources to compile the types of large, national-level datasets that the European Union, India, or the United States has. Moreover, researchers have criticized these organizations and countries for failing to target equity gaps and facilitate resources for the most marginalized populations [20,21]. In these cases, more data does not mean and has not meant more progress for the most impoverished, at-need communities around the world.

Moreover, many developing nations (defined as sovereign states with a lower Human Development Index than developed nations per the International Monetary Fund) in South America, Africa, and Asia do not report local- or national-level data beyond information shared with the OECD, rendering it difficult for developed nations, charitable non-profit organizations, and schools themselves to make data-informed decisions to improve the education and lives of children, their families, their local communities, and their nations. As a result, this study will explore how developed or developing nations can assemble large, inclusive educational datasets, using the United States as an exemplar and deeply flawed model. Although the U.S. has built enviable educational datasets, these datasets are often compiled inequitably and do not allow for appropriate disaggregation to inform targeted invention and policy work to assist children and families most in need. By learning from the U.S.—the positives and negatives—other countries can compile datasets in an equitable fashion to ensure that minoritized populations are heard and supported by their school systems and governments.

3. How Large Education Datasets Are Compiled and Equity Implications

3.1. Surveying

From the 1800s through the 1990s, the primary method that federal, state, and local governments have used to compile large education datasets is surveying. Historians estimate that early Sumerian societies around 3200 B.C. conducted censuses of their population to distribute resources and plan for levees and canals to ensure adequate water and food supply [22]. In modern societies, many countries have censuses written into their founding government documents, including the United States, Australia, several South American countries, and most of the European Union [22]. Other developing and developed countries, such as China and India, began mass collecting and publishing census data in the 1990s, and globally, nearly all forms of the census have included questions related to educational attainment level and the number of school-aged children in the household [22].

However, multiple issues arise when promoting equitable data collection for developing educational datasets through survey methods. First, organizations such as the OECD and the European Union now gather data online through Internet-based questionnaires and other methods using Internet technologies. By contrast, many developing nations do not have access to high-speed Internet—or any Internet—to facilitate effective and efficient data collection, especially in rural areas. Moreover, many developing nations have large swaths of people spread across rural, sparsely populated areas of the country, rendering robust and equitable data collection nearly impossible in countries within the Latin American and Caribbean region, Burundi, Uganda, and Nepal where rural populations comprise over 80% of the overall citizenry in each country [7].

Beyond geographic and technological limitations, many developing nations' governments do not have the human or financial resources to staff survey developers, census takers, or data architects to administer the work and disseminate its results. For instance, the United States begins its hiring process for its ten-year census two years before its administration, usually hiring over 200,000 temporary workers to complete the work [23]. Moreover, the U.S. Department of Education specifically created the National Center for Education Statistics to help liaise with schools to gather and disseminate educational data [15]. In these cases, many developing nations do not have the resources to create such offices and protocols to gather consistent, representative, reliable education data at any time interval, much less on a yearly basis as is the status quo in the United States and many other developed nations.

Finally, a wealth of education data is often tied to government funding or grant administration, requiring educational organizations to report data to their funding agency, usually a local-, state-, or federal-level entity. Although this method is not surveying in a typical sense, there are yearly reports that institutions of higher education must complete that often arrive in the form of a questionnaire. For instance, in the United States context, the process for distributing federal student aid to postsecondary students is mediated by the U.S. Department of Education through a program called Title IV, which authorizes U.S. institutions of higher education to administer financial aid programs through federal funds. As federal student aid is responsible for most of the student financing in the United States, there are over 6000 Title-IV-participating institutions of higher education in the United States. To participate, institutions must regularly report education data to the U.S. Department of Education related to the amount and type of aid that their students are receiving, as well as students' academic progress indicators [24]. Here, nationally representative educational datasets are being created in part by federal programming that requires institutional data reporting, yet many countries may not have these policy mechanisms in place through federal programs to gather such data.

3.2. Technologically Mediated Data Sharing Agreements

One of the largest international data-sharing platforms is the Statistical Data and Metadata eXchange (SDMX), sponsored by seven international organizations: the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat (Statistical

Office of the European Union), the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), the United Nations Statistical Division (UNSD), and the World Bank. SDMX is a technology and data-sharing initiative that “aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries” [25] (para. 3). Extending the survey work performed by individual nations, the SDMX allows for larger, international organizations to integrate their data into an even larger repository, allowing for unique collaborations, such as the OECD working with the International Monetary Fund, to better understand how international monetary policies may affect low-GDP nations.

However, developing nations that cannot perform the national-level survey work to lay the foundation for international data sharing thus cannot reap the benefits of international platforms such as SDMX. In this case, educational datasets across nations may be further stratified by efforts such as the ones by SDMX, with developed nations already able to gather their own national-level data in addition to reaping the benefits of international data sharing, collaboration, and joint policy development. As a result, it is critical for developed nations to scaffold the efforts of developing nations to begin the national-level survey work to allow for developing nations to participate in international data-sharing agreements, such as SDMX.

3.3. Collaborative Conglomeration Efforts

Independent researchers have also begun to integrate single-year datasets from organizations such as the OECD to compile large, longitudinal educational datasets to inform how policies and other administrative mechanisms influence the field of education over time. Barro and Lee (2013) have repeatedly conglomerated UNESCO data to compile a large educational dataset from 1950 until 2010 across 146 countries, disaggregated by sex and at five-year intervals [26]. Because of their conglomeration efforts, the researchers were able to use the data to evaluate how human capital is produced through years of schooling and the compositional education attainment of citizens. In all, the researchers found that schooling has a direct and positive impact on human capital development, and after controlling for other factors, the researchers also found that individual rate-of-return for one additional year of school was between 5 and 12% per individual [26].

At the country level, Moore’s (2022) evaluation of state-level data from two Indian states [27] and Bo et al.’s (2019) use of administrative data from China also serve as evidence that conglomerated educational datasets can drive empirical inquiry and inform policy change toward equity [28]. Moore (2022) combined datasets from two state-level datasets in India to reveal that there were large school-level effects in terms of student performance, suggesting that India’s state-level datasets could reveal state-to-state stratification that could inform Indian education policy [27]. Similarly, Bo et al. (2019) analyzed an administrative dataset from each of China’s postsecondary institutions, exploring how standardized test scores predict how students find an academic match with their institution [28]. The researchers learned that Chinese college students would reduce their probability of mismatch by 18% if they were allowed to submit their college preferences after learning their standardized test scores and not before [28]. Again, by accessing a large, national dataset in a postsecondary context, researchers were able to evaluate college matches and potentially inform national policy related to college admissions and student choice.

In 2021, State of California (USA) legislation approved funding to create a comprehensive suite called the California Cradle-to-Career Data System. This system would merge previously disconnected data systems from schools, colleges, social services agencies, financial aid providers, and employers [29]. Streamlining these data systems will allow various stakeholders to easily access information, resources, and data [30]. By using the California Cradle-to-Career Data System, students and families will be able to access pertinent information about college opportunities and other social services (e.g., medical care) in addition to formally applying to colleges and financial aid. Educators, on the other hand,

will have a centralized platform to monitor the progress and completion of college and financial aid applications. This is essential to building equity because it gives educators the ability to provide targeted support to under-resourced areas or specific communities of people. Lastly, for policymakers, researchers, and advocates, this comprehensive system will provide longitudinal student and employment outcome data that will allow them to see trends and inform potential interventions [31]. While still in its planning stage, the Cradle-to-Career Data System provides a glimpse into the future in terms of how multiple data systems across sectors can be streamlined into one cross-sector data system that provides information and data to various stakeholders to promote equity and inform change.

3.4. Collaborative Comparative Efforts

Understanding the parameters of higher education in international contexts is necessary to make sense of how institutional data are developed and used. While sociopolitical and cultural variability exists in countries around the world, data are becoming more prominent in higher education institutions. From here, institutions and countries may be able to collaborate to compare data and seek out equity-based solutions to interinstitutional or intercontinental problems.

It is important to note that although countries have generated institutional data, the contexts in which they use this data vary; for instance, in Europe and Asia, these institutional data systems are focused on public policies [21]. According to Lepori et al. (2022), higher education in Asia is often compared to examples from China and Thailand in that this type of training and education is extensive and diverse. There has been an increase in higher education in Asia and as such, there is more emphasis on discussing institutional data [21].

Lepori et al. (2022) indicated that the institutional data in the United States, Europe, and Asia contain similar information that is required by the state (education and existing higher education resources), and these data are equally important as the data from UNESCO and OECD [21]. This is indicative of the need for collaborative comparative efforts to move forward in the development of appropriate and relevant institutional data for high education.

4. Limitations

Yet, developing nations without the human or financial capital to gather data and conglomerate datasets will remain behind developed nations. As a result, developing nations need to prioritize widespread survey administration to build local and national datasets to be able to engage with larger, internationalized datasets, thus joining the global data community to use data to make informed decisions regarding education policy and practice. First, however, governments need to inventory their current data collection procedures and consolidate efforts to begin working toward robust, longitudinal datasets. Then, as developing nations are generating the capacity to perform this survey work, researchers and policymakers in these countries could begin to learn how other countries use technologies such as SDMX to explore how their own country could utilize and benefit from such a resource. Finally, educational leaders need to engage with these data to make equitable decisions and allocate resources to the most marginalized communities, rather than merely collecting and reporting on the data.

5. Equity Issues Related to Survey Instruments and Data Collection

As mentioned above, there are five primary hurdles to developing survey instruments and mass-collecting data in developing countries:

- (1) Human capacity: Who will develop instruments and gather data? Does an organization have the human capability to develop data collection instruments and carry out the work?
- (2) Financial capacity: who will finance the data collection efforts?

- (3) Technological capacity: are there technological resources available to render the data collection process more efficient and effective?
- (4) Geography: are all areas of the country physically accessible without considerable resource allocation, and do countries know where their people are?
- (5) Sociopolitical contexts and variability: Certain organizations and countries are situated within sociopolitical contexts that may not be amenable to truly equitable data collection and database building. For instance, some countries openly oppress and discriminate against queer people [32], whereas in other countries people who identify as women are not allowed to attend school or enjoy various social freedoms that women enjoy in different countries [33,34]. As a result, many countries and organizations may not be willing or able to gather accurate data for equity for all people.

Once developing nations have negotiated these five hurdles, it becomes crucial that initial or current survey instruments and data collection strategies are built with equity in mind. Robust datasets can be powerful tools for educational stakeholders; however, depending on how robust the data collection is and what variables were included within data collection instruments can either strengthen or weaken its utility. One way for datasets to become more robust is to expand the survey instrument to gather specific demographic characteristics beyond what is typically gathered. Most survey instruments meant to capture educational data include demographic questions about a respondent's race and/or ethnicity, gender identity and/or sexual orientation, religion, and other salient identities. However, many survey instruments deployed by the most developed countries, such as the United States, do not adequately specify groups of people, especially given the long history of immigration to the United States from countries around the world. By including more specific questions about the participant's identities, the dataset allows users to explore potential trends within and between groups. Whether the differences are stark or nuanced, the ability for users to compare and contrast data between and within groups allows for better data analysis.

The following sections will detail three examples of why it is important to expand the questions about participants' identities. While not exhaustive of all identities, we highlight examples of how gathering specific, accurate information on participants' identities is critical to advancing equity in education through large datasets.

5.1. The Importance of Expanding the Race/Ethnicity Variable for People of Color

In no uncertain terms, homogenized racial and ethnic categories do little to help understand cultural nuances between and among different races and ethnicities. A crucial example of the problematic nature of how the U.S. has gathered educational data related to race and ethnicity is the current situation facing Asian Americans. Within the racial fabric in the United States, Asian Americans find themselves in a peculiar position [35]. Beginning in the 1960s, Asian Americans, who were once viewed as a threat to White Americans regarding jobs and sheer numbers in specific regions, became the model minority due to their quiet demeanor, work ethic, and educational prowess [36]. However, inequitable data collection initiatives have grouped Asian Americans into an inauthentic homogenized group that does not allow for pointed, accurate data analysis and disaggregation by race or ethnicity.

To begin with, many Westernized data collection instruments do not gather race or ethnic data beyond homogenized categories, typically including White and/or Caucasian, Black and/or African American, Hispanic and/or Latinx, and Asian American. These categories are problematic, as researchers have articulated many equity gaps between racial and ethnic groups within these broader categories [37,38]. For instance, in Western contexts, especially the United States, the model minority myth portends that "Asian Americans achieve universal and unparalleled academic and occupational success" which "perpetuate ignorance and distorted perceptions of the realities that this population" faces [37] (p. 6).

Here, the way in which researchers and other stakeholders gather racial and ethnic data may perpetuate the model minority myth, especially as it relates to Asian American

educational achievement data. For instance, the United States' National Center for Education Statistics recently published disaggregated statistics regarding Asian American postsecondary success, finding that 54% of Asian American adults aged 25 or older held a bachelor's degree or higher. However, when parsed by ethnic group, these achievement data reveal that 74% of Asian Indians aged 25 or older held a bachelor's degree or higher, while Cambodian (16%), Hmong (18%), Laotian (18%), Burmese (21%), and Vietnamese (29%) adults have different levels of education experience [39]. Here, researchers in developing countries must build survey instruments that allow for survey respondents of color to narrowly define their racial and/or ethnic group to best represent the population and allow policymakers to allocate resources equitably given educational access and success gaps.

5.2. The Importance of Expanding the Gender Variable for the Queer Community

The same disaggregation that must occur within racial and/or ethnic groups should also occur within gender identities if feasible given the cultural context, as researchers must move behind the gender binary and allow queer survey respondents to accurately and narrowly define their own gender identity. However, we already addressed issues in several countries where some countries openly oppress and discriminate against queer people [32]. In these circumstances, it may be difficult or impossible to gather truly equitable and democratic datasets where everyone's voice—and their personally accurate identities—is captured accurately and in a culturally-responsive way. For decades, scholars of queer studies have criticized the male–female sexuality binary and man–woman gender binary of data collection and analysis, insisting that people who do not feel that one of the binary categories describes them have felt their sense of existence silenced and marginalized [40,41].

Research regarding the higher education experiences of transgender people has suggested that people who do not view school supports as gender neutral, such as gender-inclusive bathrooms and nondiscrimination policies that are inclusive of diverse and non-binary gender identities, may self-exclude from higher education, limiting the educational opportunities for non-binary conforming individuals [40,41]. As a result, researchers and social justice advocates in developing nations must first challenge oppressive societal norms, such as the subjugation of queer people, and work to facilitate more welcoming, inclusive societies on the basis of gender identity. Then, researchers and policymakers should build survey instruments that allow non-binary conforming respondents to narrowly define their gender identity to best represent the population and allow policymakers to allocate resources equitably.

5.3. Intersectional Analysis: Gender and Race and/or Ethnicity

Collecting and disaggregating data beyond gender identity and race and/or ethnicity has been found to be critical for intersectional education equity. For instance, in the United States context at the postsecondary level, men outpaced women in college access and bachelor's degree attainment from the inception of U.S. higher education in the 1600s until roughly the year 2000. Around 2000, women surpassed men in both college access and completion, with 10% more women earning bachelor's degrees than men [42]. Now, in both the U.S. context and around the world, women comprised roughly 60% of the overall postsecondary enrollment in the United States in 2021 [43], and recent research suggested that men, across at least 18 other countries, are less likely to access and complete their K-12 and higher education than women, continuing the global trend of inequitable education gaps between men and women [44].

However, integrating both gender and race and/or ethnicity into data collection and analysis reveals even starker, more critical equity gaps. For instance, Sáenz and Ponjuan (2008) highlighted the improvement that Latinx college students had made in accessing U.S. higher education over prior decades, yet Latinx men had the lowest high school graduation rates of men across all ethnic groups [45]. These researchers also found that over 60% of postsecondary credentials were earned by Latinx women [45]. After analyzing

large educational datasets, Sáenz et al. (2015) set out to fill these educational equity gaps by establishing Project MALES, a research-to-practice mentoring program that provides specific mentoring and education interventions for young men of color to improve their access to and success within schools at both the secondary and postsecondary level [46].

5.4. *Disproportionality in Education for People of Color and People with Disabilities*

The issue of disproportionality in special education has persisted over the decades. In the United States, the disproportionate representation of students of color and with disabilities in special education continues to be reported and studied in the literature. Disproportionality is referred to as “the overrepresentation and underrepresentation of a specific demographic group in special education relative to the presence of this group in the overall student population” [47] (p. 1). There has been perpetual rampant discrimination against culturally and linguistically diverse students and students with disabilities, thereby resulting in disproportionality. While in the United States, the Individuals with Disabilities Education Act has been amended to account for this challenge of misrepresentation of minority groups in special education, more national data are needed to further address this grave problem [47]. Overall, emphasis needs to be placed on gathering demographic data on race and ethnicity and disability status, allowing an intersectional view of oppression, and ultimately, equity.

According to Artiles and Trent (1994), researchers have yet to create thorough analyses incorporating both history and social factors in special education issues that would assist in the development of better policies and practices for marginalized groups [48]. It is critical to consider these factors when collecting national, and large-scale data, as Van Roekel (2008) proposed in a call to action for policymakers and other stakeholders to collaborate in reducing disproportionality in special education [47]. However, the reality is that there needs to be richer data that shed light on this issue. While it is well-documented that disproportionality exists in the United States [47,48], little is known about this issue in other countries around the world, suggesting that developing nations could prioritize this work to inform equitable policies for people with disabilities [49].

5.5. *Other Critical Variables: Income Status, Educational Attainment, and Family Structure*

Beyond gathering expanded categorical data related to race and/or ethnicity, gender identity, and disability status, there are several other critical demographics for researchers to integrate into large educational datasets to improve equitable outcomes for students, their families, and their communities. First, and at all levels, it is critical to gather the household or family income level as a proxy of a student’s socioeconomic status. After decades of analysis of large, longitudinal educational datasets, it has been established that understanding an individual’s, school’s, or community’s socioeconomic status can help identify gaps in educational services for low-income people [16,50], including high-quality teachers and school buildings, transportation to school, meals at school, and other factors known to affect one’s educational experiences and outcomes.

Although closely related to income status, it is also critical for education researchers and policymakers to gather educational attainment data at all levels, including access and completion rates at the intermediate/middle school, secondary/high school, and postsecondary/higher education levels. In most developed nations, educational attainment data has been used to understand how or if people have equitable access to educational institutions and whether policies can positively impact one’s ability to earn secondary and postsecondary credentials and improve their economic future [3,4,9,16]. At the postsecondary level, a wealth of research has emerged from large, national datasets to suggest that students who do not have parents who have earned a bachelor’s degree (known as first-generation college students) do not access postsecondary education or earn salaries at the level of their peers [51]. Moreover, equity gaps widen between first-generation students of color and White peers, suggesting that it is important to understand a student’s race and/or ethnicity and their parent’s educational attainment to identify and stem equity

gaps [51]. As a result, it is critical to gather data related to educational attainment to build datasets that understand education gaps at multiple levels to advocate for policy to fill these gaps.

Finally, it is critical to understand the family or household structure that a student is raised in to understand that student's educational opportunities and future. Researchers across the world have investigated the roles that being raised in foster care [52], being a child of divorced parents [53], being adopted [54], and growing up in diverse living environments play in the educational attainment of children and young adults, often finding that children without supportive and consistent parenting and mentorship will not have access to the same educational experiences as privileged peers [52–54]. Although the national census in many countries may capture family or household information [55], that information must be synthesized with educational data to best understand how and where a student is raised and whether they face educational hurdles or lack opportunity.

6. Conclusions

Ultimately, developed nations have provided guidance for developing nations when building large educational datasets to improve educational decision-making and fill equity gaps. As most developed nations have done, central governments in developing nations should continue to build both human and financial capacities to survey their population and pay close attention to demographic information that has been found to impact students' educational success and economic development. This implies broad, equitable surveying of diverse geographic areas to ensure that all people are counted, and their data contributed to the local- or national-level dataset. Here, developing nations will likely need to develop relationships with community-based organizations and collaborate with local communities to understand how to survey the people and understand local demographics. Building trust and communicating clearly with local communities could help ensure that surveying is robust and accurate, as well as ensure that resources can be distributed equitably once data are collected and analyzed.

Moreover, researchers should develop survey instruments that capture a wide range of races and/or ethnicities, gender identities, disability statuses, and other personal demographics to ensure people are accurately and authentically counted and supported. As robust as their data are, the shortcomings of the U.S. and E.U. datasets are that demographics are often reported by categories that are far too large and miss the nuance that is required to provide targeted educational interventions. Whether survey instruments are newly developed or new iterations of old designs, researchers should expand demographic categories to better understand—and respect—all people to improve their educational opportunities and outcomes.

In general, developing nations must consider the five major limitations of building large educational datasets: human capacity, financial capacity, technological capacity, geography, and sociopolitical contexts and variability. Although developing nations such as India and developed nations such as the United States, the European Union, Australia, and China have enviable datasets and educational resources, other developing nations can follow their lead and begin developing inclusive survey instruments and collaborating with communities to build rich datasets capable of being integrated with international data exchanges, such as SDMX. In modern society, forging a path toward educational equity will require data-driven decision-making to uncover equity gaps and distribute resources equitably, and developing nations can serve their people through the equitable building of educational datasets to improve lives everywhere.

Author Contributions: Conceptualization, Z.W.T.; methodology, Z.W.T.; formal analysis, Z.W.T., J.K., C.C. and J.C.; investigation, Z.W.T., J.K., C.C. and J.C.; resources, Z.W.T., J.K., C.C. and J.C.; data curation, Z.T, J.K., C.C. and J.C.; writing—original draft preparation, Z.T, J.K., C.C. and J.C.; writing—review and editing, Z.W.T., J.K., C.C. and J.C.; visualization, Z.W.T., J.K., C.C. and J.C.; supervision, Z.T, J.K., C.C. and J.C.; project administration, Z.W.T., J.K., C.C. and J.C.; funding acquisition, Z.W.T. All authors have read and agreed to the published version of the manuscript.

Funding: This Special Issue Was Funded by the International Joint Research Project “Education Development and Social Justice”, Faculty of Education, Beijing Normal University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mandinach, E.B. A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educ. Psychol.* **2012**, *47*, 71–85. [CrossRef]
- Schildkamp, K.; Poortman, C.; Luyten, H.; Ebbeler, J. Factors promoting and hindering data-based decision making in schools. *Sch. Eff. Sch. Improv.* **2016**, *28*, 242–258. [CrossRef]
- Volante, L.; Ritzen, J. The European Union, education governance and international education surveys. *Policy Futur. Educ.* **2016**, *14*, 988–1004. [CrossRef]
- Hakhverdian, A.; van Elsas, E.; van der Brug, W.; Kuhn, T. Euroscepticism and education: A longitudinal study of 12 EU member states, 1973–2010. *Eur. Union Politics* **2013**, *14*, 522–541. [CrossRef]
- Mandinach, E.B.; Gummer, E. Data-Driven Decision Making: Components of the Enculturation of Data Use in Education. *Teach. Coll. Rec.* **2015**, *117*, 1–8. [CrossRef]
- Organisation for Economic Co-Operation and Development [OECD]. Education: Online Education Database. 2022. Available online: <https://www.oecd.org/education/database.htm> (accessed on 15 January 2023).
- Organisation for Economic Co-operation and Development [OECD]. Education. 2022. Available online: <https://www.oecd.org/education/> (accessed on 13 January 2023).
- Organisation for Economic Co-Operation and Development [OECD]. Graduation Rates and Entry Rates. 2022. Available online: https://stats.oecd.org/Index.aspx?datasetcode=EAG_GRAD_ENTR_RATES (accessed on 13 January 2023).
- European Commission. *European Education Area: Quality Education and Training for All*; European Commission: Luxembourg, 2022. Available online: <https://education.ec.europa.eu/> (accessed on 13 January 2023).
- National Informatics Centre. Statistics, New. Ministry of Education, Government of India. 2022. Available online: https://www.education.gov.in/en/statistics-new?shs_term_node_tid_depth=All&page=5 (accessed on 13 January 2023).
- World Bank Group. *Government Expenditure on Education, Total (% of GDP)—India*; World Bank: Bretton Woods, NH, USA, 2022. Available online: <https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?locations=IN> (accessed on 13 January 2023).
- National Informatics Centre. Statistics of School Education, 2008–09. Ministry of Education, Government of India. 2022. Available online: https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-School_2008-09.pdf (accessed on 13 January 2023).
- Sharma, S. Will Budget 2022 Heal India’s Sagging Education System? *India Today*. 2022. Available online: <https://www.indiatoday.in/diu/story/budget-india-education-system-covid-school-closures-1906438-2022-01-30> (accessed on 13 January 2023).
- ASER Centre. Annual Status of Education Report (Rural). ASER Centre. 2021. Available online: <http://img.asercentre.org/docs/aser2021forweb.pdf> (accessed on 15 January 2023).
- National Center for Education Statistics. *Federal Education Data Collection: Celebrating 150 Years*; National Center for Education Statistics: Washington, DC, USA, 2017. Available online: https://nces.ed.gov/surveys/annualreports/pdf/Fed_Ed_Data_Collection_Celebrating_150_Years.pdf (accessed on 15 January 2023).
- Goldin, H. *NBER Working Paper Series On Historical Factors in Long Run Growth: A Brief History of Education in the United States*; National Bureau of Economic Research: Cambridge, MA, USA, 1999. Available online: https://www.nber.org/system/files/working_papers/h0119/h0119.pdf (accessed on 15 January 2023).
- Institute of Education Sciences. An Overview of NAEP. National Center for Education Statistics. 2021. Available online: https://nces.ed.gov/nationsreportcard/subject/about/pdf/naep_overview_brochure_2021.pdf (accessed on 15 January 2023).
- United States Government. National Education Goals Panel. Federal Register. 1990. Available online: <https://www.federalregister.gov/agencies/national-education-goals-panel> (accessed on 15 January 2023).
- United States Department of Education. Office of the Inspector General: American Recovery and Reinvestment Act. U.S. Department of Education. 2019. Available online: <https://www2.ed.gov/about/offices/list/oig/recoveryact.html> (accessed on 15 January 2023).
- Data Quality Campaign. *Using Educational Data for Equity*; The Data Quality Campaign: Washington, DC, USA, 2023. Available online: <https://dataqualitycampaign.org/our-work/policy-areas/data-for-equity/> (accessed on 15 January 2023).
- Lepori, B.; Borden, V.M.H.; Coates, H. Opportunities and challenges for international institutional data comparisons. *Eur. J. High. Educ.* **2022**, *12*, 373–389. [CrossRef]
- Whitby, A. *The Sum of the People: How the Census has Shaped Nations, from the Ancient World to the Modern Age*; Basic Books: New York, NY, USA, 2020.

23. Federal Student Aid. *Knowledge Center Home: Overview of Title IV*; U.S. Department of Education: Washington, DC, USA, 2022. Available online: <https://fsapartners.ed.gov/knowledge-center/library/functional-area/Overview%20of%20Title%20IV> (accessed on 15 January 2023).
24. United States Bureau of Labor Statistics. *Beyond the numbers: How Did the 2020 Census Affect Employment?* United States Government: Washington, DC, USA, 2022. Available online: <https://www.bls.gov/opus/btn/volume-11/how-did-the-2020-census-affect-employment> (accessed on 15 January 2023).
25. Statistical Data and Metadata eXchange. What Is SDMX? SDMX. 2022. Available online: https://sdmx.org/?page_id=3425 (accessed on 18 January 2023).
26. Barro, R.J.; Lee, J.W. A new data set of educational attainment in the world, 1950–2010. *J. Dev. Econ.* **2013**, *104*, 184–198. [CrossRef]
27. Moore, R. Variation, context, and inequality: Comparing models of school effectiveness in two states in India. *Sch. Eff. Sch. Improv.* **2022**, *33*, 588–609. [CrossRef]
28. Bo, S.; Liu, J.; Shiu, J.-L.; Song, Y.; Zhou, S. Admission mechanisms and the mismatch between colleges and students: Evidence from a large administrative dataset from China. *Econ. Educ. Rev.* **2018**, *68*, 27–37. [CrossRef]
29. State of California. About the Cradle-to-Career System. State of California Cradle-to-Career System. 2023. Available online: <https://c2c.ca.gov/about-cradle-to-career-system/> (accessed on 18 January 2023).
30. WestEd. California Cradle-to-Career Data System. California All: California Data System. 2023. Available online: <https://cadatasystem.wested.org/> (accessed on 22 January 2023).
31. California Data System. About the Cradle to Career System. The Cradle to Career Data System. 2023. Available online: <https://c2c.ca.gov/about-cradle-to-career-system/> (accessed on 21 January 2023).
32. Yavuz, M.; Byrne, S. Center for Peace and Justice Education at Villanova University Violence Against the Queer Community in Turkey. *J. Peace Justice Stud.* **2021**, *30*, 102–125. [CrossRef]
33. Baldwin, E.; Fernea, E.W. Children in the Muslim Middle East. *J. R. Anthr. Inst.* **1998**, *4*, 145–146. [CrossRef]
34. Hamdan, A. Women and education in Saudi Arabia: Challenges and achievements. *Int. Educ. J.* **2005**, *6*, 42–64.
35. Hartlep, N.D.; Porfilio, B.J. *Killing the Model Minority Stereotype: Asian American Counterstories and Complicity*; Information Age Publishing: Charlotte, NC, USA, 2015.
36. Au, W. Asian American racialization, racial capitalism, and the threat of the model minority. *Rev. Educ. Pedagog. Cult. Stud.* **2022**, *44*, 185–209. [CrossRef]
37. Museus, S.D.; Kiang, P.N. Deconstructing the model minority myth and how it contributes to the invisible minority reality in higher education research. *New Dir. Inst. Res.* **2009**, *2009*, 5–15. [CrossRef]
38. Poon, O.; Squire, D.; Kodama, C.; Byrd, A.; Chan, J.; Manzano, L.; Furr, S.; Bishundat, D. A Critical Review of the Model Minority Myth in Selected Literature on Asian Americans and Pacific Islanders in Higher Education. *Rev. Educ. Res.* **2016**, *86*, 469–502. [CrossRef]
39. National Center for Education Statistics. *Status and Trends in the Education of Racial and Ethnic Groups 2018*; National Center for Education Statistics: Washington, DC, USA, 2019. Available online: <https://nces.ed.gov/pubs2019/2019038.pdf> (accessed on 18 January 2023).
40. Goldberg, A.E.; Beemyn, G.; Smith, J.Z. What Is Needed, What Is Valued: Trans Students’ Perspectives on Trans-Inclusive Policies and Practices in Higher Education. *J. Educ. Psychol. Consult.* **2018**, *29*, 27–67. [CrossRef]
41. Jones, A. *Trans Theory and Gender Identity in Education*; Oxford University Press: Oxford, UK, 2021. [CrossRef]
42. Parker, K. *What’s Behind the Growing Gap between Men and Women in College Completion?* Pew Research Center: Washington, DC, USA, 2021. Available online: <https://www.pewresearch.org/fact-tank/2021/11/08/whats-behind-the-growing-gap-between-men-and-women-in-college-completion/> (accessed on 18 January 2023).
43. Stewart, M. The Enrollment Crisis for Men Continues to Worsen. Insight into Diversity. 2022. Available online: <https://www.insightintodiversity.com/the-enrollment-crisis-for-men-continues-to-worsen/> (accessed on 17 January 2023).
44. Stoet, G.; Geary, D.C. Gender differences in the pathways to higher education. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 14073–14076. [CrossRef]
45. Saenz, V.B.; Ponjuan, L. The Vanishing Latino Male in Higher Education. *J. Hisp. High. Educ.* **2008**, *8*, 54–89. [CrossRef]
46. Sáenz, V.B.; Ponjuan, L.; Segovia, J.; Viramontes, J.D.R. Developing a Latino Mentoring Program: Project MALES (Mentoring to Achieve Latino Educational Success). *New Dir. High. Educ.* **2015**, *2015*, 75–85. [CrossRef]
47. Van Roekel, N.P.D. *Disproportionality: Inappropriate Identification of Culturally and Linguistically Diverse Children*; National Education Association: Washington, DC, USA, 2008. Available online: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.3895> (accessed on 17 January 2023).
48. Artiles, A.J.; Trent, S.C. Overrepresentation of Minority Students in Special Education. *J. Spéc. Educ.* **1994**, *27*, 410–437. [CrossRef]
49. Charran, C.; Sorrells, A.M.; Cooc, N. Quantitative Research Methods and Design for Investigating Inclusive Education in the Caribbean. In *Achieving Inclusive Education in the Caribbean and Beyond*; Springer: Berlin, Germany, 2019; pp. 51–65.
50. Meschi, E.; Scervini, F. A new dataset on educational inequality. *Empir. Econ.* **2014**, *47*, 695–716. [CrossRef]
51. Trevino, N.N.; DeFreitas, S.C. The relationship between intrinsic motivation and academic achievement for first generation Latino college students. *Soc. Psychol. Educ.* **2014**, *17*, 293–306. [CrossRef]
52. I Männistö, I.; A Pirttimaa, R. A review of interventions to support the educational attainments of children and adolescents in foster care. *Adopt. Foster.* **2018**, *42*, 266–281. [CrossRef]
53. Amato, P.R. Research on Divorce: Continuing Trends and New Developments. *J. Marriage Fam.* **2010**, *72*, 650–666. [CrossRef]

54. Barclay, K.J. Birth order and educational attainment: Evidence from fully adopted sibling groups. *Intelligence* **2015**, *48*, 109–122. [[CrossRef](#)]
55. United States Census Bureau. *CPS historical Time Series Tables*; U.S. Census Bureau: Suitland, MD, USA, 2022. Available online: <https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html> (accessed on 17 January 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.