education sciences

Article

Using Big Data for Educational Decisions: Lessons from the Literature for Developing Nations

Zach W. Taylor [1,*], Chelseaia Charran [2] and Joshua Childs [3]

[1] Education and Human Sciences, The University of Southern Mississippi, Hattiesburg, MS 39406, USA
[2] Département de Psychoéducation, Université du Québec à Trois-Rivières, Trois-Rivières, QC G9A 5H7, Canada
[3] Department of Educational Leadership and Policy, The University of Texas at Austin, Austin, TX 78712, USA
* Correspondence: z.w.taylor@usm.edu

Abstract: Educational leaders from developing countries may be tasked with using big data to help inform educational decisions. Although many researchers have explored how to use big data or datasets to help solve educational problems, few studies have articulated how educational researchers and leaders from developing nations can use big data to make educational decisions. This study provides a literature review and takes a position to help educational leaders from developing nations use big data to make educational decisions and understand the strengths and weaknesses of using data to drive decision making. Moreover, this study addresses how datasets may be limited and how educational leaders can understand these limitations when using big data.

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

check for updates

Citation: Taylor, Z.W.; Charran, C.; Childs, J. Using Big Data for Educational Decisions: Lessons from the Literature for Developing Nations. *Educ. Sci.* **2023**, *13*, 439. https://doi.org/10.3390/educsci13050439

Academic Editor: Han Reichgelt

Received: 20 February 2023
Revised: 17 April 2023
Accepted: 22 April 2023
Published: 25 April 2023

1. Introduction

Although its definition continues to change as the technological ecosystem changes, big data can be referred to as "data that is so large, fast or complex that it's difficult or impossible to process using traditional methods" [1] (para. 1). Famed data analytics guru Doug Laney conceptualized big data into three Vs: volume, velocity, and variety. Volume refers to how organizations gather data from a variety of sources, including computers, smart devices, cameras, social media platforms, and many others. In education settings, educators often gather data from student interactions with curricular materials across many of these sources, not to mention the data gathered by educational leaders at the school, district, region, or national level. The second V, velocity, refers to the growth of the Internet's integration with everyday devices and processes, such as e-books with embedded Internet resources. The velocity of big data requires organizations to be nimble and flexible, as data can be captured—or lost—at unprecedented speed, if the organization has adequate data collection and storage capacity. Finally, the third V—variety—implies that data can come in many different formats, from traditional databases of information in columns and rows to highly disorganized and unstructured data [2], such as multimedia, global positioning system (GPS) information, or Microsoft PowerPoint files. This variety places educational organizations in difficult positions, as educational organizations are resistant to change given their bureaucratic nature, with many organizations only able to analyze more traditional data in traditional ways.

Despite the traditional nature of education, the era of big data has arrived in the field of education on a global scale. As the Internet became widely available to educational organizations in the 1990s and online education has exploded in growth and popularity since 2000, many educational leaders and policymakers now have access to more data than ever before [3]. As a result, both governments and educational organizations have made considerable efforts to use large datasets to make educational decisions, including those related to curriculum and instruction, program development, policy advocating, resource allocation, and countless other educational decisions [4–7].

However, as data is continuously created, collected, and analyzed by educational researchers, those collecting it may reach a point of diminishing returns: How much data is too much? And in an era where nearly everything can be observed and digitally documented, when do educators reach a point of data exhaustion and overload? In a discussion of big data in business circles, data analysts often say you cannot manage what you do not measure [8], but surely the inverse is also true: You cannot measure what you cannot manage. Educational leaders and policymakers face this challenge in many aspects of their operations, as many educational systems in developing nations may be in their nascent stages of conceptualizing data collection, much less engaging with big data analytics.

Moreover, the utility of big data requires both a technical expertise and a level of communication that many bureaucratic educational organizations simply may not possess, especially in under-resourced developing nations. Here, educational organizations in developing nations are placed in an interesting position, as data is now more available than ever from a wide variety of sources and stakeholders, and this data could prove transformative in the efforts that organizations make to become more efficient and effective for their students. However, developing nations with limited human and financial capital, complex bureaucratic organizations, and limited technical capacity may need to catch up to the speed in which big data has advanced and will continue to do so.

Additionally, the arrival of neoliberal policies and agendas in many developing nations has placed educational organizations in difficult positions regarding the country- and local-level allocation of resources, both human and financial. Core tenets of neoliberalism—including the privatization of public functions, the deregulation of industry, and reductions in spending on public initiatives—has been felt in educational contexts within developed and developing nations. As a result, many developing nations may have limited human and financial resources than their peers, with this stratification and inequity exacerbated by neoliberal policies enacted by national or local governments.

From here, this discourse provides an overview of how educational organizations have strategically utilized and benefited from data-driven decision-making using large educational data sets. Additionally, this review will outline several drawbacks and ethical concerns of using big data sets in education, including how uncertainties in human and financial capacity as well as limited technological capability may hinder developing nations who desire big data to make decisions but are not in the position to do so. Furthermore, accountability systems at the national, state, and local levels within developed nations have become more technologically advanced as data continues to become increasingly available and abundant. Subsequently, educational leaders and policymakers from developing nations must understand how big data sets have been used in the past and how these leaders can develop the organizational capacity to use the data to improve the lives of students and the communities to which they belong.

2. Benefits of Big Data in Education

Research has suggested that the appropriate engagement with big data can be a force for educational equity, evidenced by countless studies where educational stakeholders have engaged with complex data sets to identify equity gaps and improve teaching and learning and student outcomes [3–7,9–14]. The benefits of big data utility are multifarious, and the following sections will outline several crucial benefits for developing nations seeking the ability to make large-scale data-informed decisions.

2.1. Individualization through Data-Informed Teaching and Learning

As many researchers argue, the primary function of education is teaching and learning, and many scholars have pointed to the utility of big data as a driver of the improvement of teaching and learning at all levels of education. Schildkamp et al. reasoned that as schools are increasingly held more accountable for student learning, countless school districts in developed nations have used local or regional data sets to improve the manner in which students are oriented with curricular materials and how teachers are prepared for the classroom by post-secondary institutions [13]. Nazarenko and Khronusova explained that there are "incredible opportunities for individualization and personalization of the student's path to content mastery based on adaptive learning or competency-based education" [15] (p. 676), as schools in developed nations often have access to increasingly technologically advanced modes of content delivery, and thus access to even more data to make even more decisions.

Moreover, Nazarenko and Khronusova have explained that teachers and administrators would likely have ample data to target educational inequities, such as the challenges faced by students with disabilities. Here, the authors reasoned that schools could provide "targeted interventions to improve student's success and to reduce overall costs to students and institutions" [15] (p. 676). In the Australian context, the national government has engaged with big data analytics to provide teachers and administrators with information to personalize learning to align with national policies related to teacher and school effectiveness, including the stemming of educational inequities [16].

Wang went into further detail, explaining that schools can move far beyond "student demographics, test scores, and psychological questionnaires" toward more fine-grained data collection methods, such as "computer mouse clicks, number of attempts, learning browsing patterns, online chats, discussion forum participations, and visual and facial reactions" [3] (p. 382). Although these approaches require technologically mediated education, the technology exists to equip classrooms with cameras and tracking devices to allow teachers to understand when students are on task and how efficient and effective their teaching style is for diverse learners [3]. Furthermore, the expansion of mobile devices and "bring your own device" initiatives has greatly expanded the walls of the classroom, allowing schools to understand not only which technologies are best for student learning and teaching by faculty, but administrators can also understand which type of hardware is most conducive to effective teaching and learning [3,6]. This insight can be facilitated by capturing and analyzing big data to inform a wide variety of teaching and learning subjects such as student attention, teacher effectiveness, relationship development, assessment types and strategies, and a plethora of others.

2.2. Broader Generalizability

Big data can also facilitate opportunities for the cross-organizational analysis of educational functions, as many researchers have suggested that big data allows for greater generalizability so that other organizations can learn from each other, especially if these organizations serve similar populations in similar geographic areas [3,17].

For instance, Crossley specifically spoke about how data can be transferred internationally to allow educational research in one nation to inform the policies and practices of education in another nation. This can be an especially important technique for developing nations where human and financial resources are limited. As Crossley explained, "With references to my own work in Kenya and Tanzania ... carried out by African researchers, perhaps in partnership with international colleagues, has much to offer, if a greater proportion of educational reform initiatives are to be translated into successful practice" [17] (p. 22). In this instance, developing nations were able to learn from each other's big data and recapture limited human and financial resources related to big data capture and analysis.

Wang also spoke to the nature of big data as facilitating generalizability, as they argued that big data often informs educational policy through mass communication over websites and through social media. As an emerging form of big data, educational policymakers can now understand public sentiment and access trend-related data to best understand how students, teachers, and other stakeholders feel toward educational policies or identify educational inequities [3]. Wang argued that this form of big data allows for generalizability, as the internet and communication technologies allow many different stakeholders to have a public-facing voice on issues facing educational institutions [3]. Although beyond the traditional student demographics and test scores to inform policy, Wang suggested that innovative and new forms of communication can allow for educational leaders to analyze big data to generalize public sentiment and inform educational policy toward equity [3].

2.3. Accountability and Measurement

Many regions and developing nations often lean on big data for accountability and measurement purposes [6,13]. At the higher education level, Macfadyen et al. reasoned that, "in the complex systems of higher education, current performance assessment and accountability policies may be the forces driving the continued focus on high–stakes testing as a means of producing comparative institutional data, despite the well–articulated weakness of such an approach for understanding student learning" [18] (p. 18). Here, although the authors point to perhaps an over-reliance on big data, many institutions of higher education often tie big data to assessment and accountability policies, for better or worse.

Likewise, Schildkamp et al. reasoned that big data allows teachers and administrators to review and confirm that they are measuring student learning, tying that learning to educational objectives and measurements, and demonstrating accountability to local, state, or national mandates and policies, many of which may be tied to important sources of educational funding [13]. Big data also gives teachers and administrators insight into current practices to improve their accountability toward educational policies, in turn allowing for educational leaders to provide educational interventions for students and support services for educators to improve the overall education system [13]. For instance, Kraft et al. analyzed administrative data from New York City school districts to learn that school safety and academic expectations were associated with lower levels of teacher turnover and higher levels of student achievement, suggesting that individual school data may be nuanced, but when combined with larger data sets, policy decisions can be made easier and in more generalizable terms [10]. These authors all emphasize the point that educational leaders need to first have well-defined goals and data available to track progress toward those goals, rendering it incredibly important for educational leaders to either be adept data managers and analysts, or to employ a team who can perform data management and analysis tasks to inform leadership [10,13].

2.4. Strategic Budget Allocation

Schildkamp et al. focused on how big data and data-driven decision-making can also inform budgetary decisions, especially on a large scale. As many national governments often disseminate resources from the national level to the regional or local level, it is critical that governments and school districts access data and explore equity gaps to disseminate funds and improve schools and communities in low-income areas [13]. Studies related to teacher turnover have found that some school districts may need to allocate budgetary resources to recruit and retain high-quality teachers, an insight only gleaned from the analysis of a large administrative data set in one of the most populous cities in the world, New York [10].

Additionally, the European Commission also gathers data from many E.U. member nations to inform how developed and developing nations can allocate budgetary resources to provide educational interventions for teachers and students, as well as understand where education systems need to be developed in both populous urban areas or rural areas [9]. Of the European Commission's strategic goals, E.U. member nations have shared data to arrive at literacy goals in primary and secondary schools and post-secondary achievement goals that have allowed individual institutions and nations to strategically allocate funds to support those initiatives [9]. Crossley's transnational work also speaks to the European Commission, as many E.U. member nations have seen the benefit of big data sharing agreements to better allocate financial resources and improve student outcomes at multiple levels [17].

3. Drawbacks of Big Data in Education

As there are countless benefits to capturing and analyzing big data and the educational context, developing nations should be particularly concerned with the many drawbacks with regard to big data and education. As many developing nations have limited resources, both human and financial, it is critical to understand the type and sculpt of big data that would best serve a particular region or an entire developing nation. As many big data initiatives take years or decades to launch, developing nations should heed these warnings as they relate to big data and educational decision-making.

3.1. Size and Overwhelm Paralysis

Even though educators should be able to make better decisions with more data, two of the three vs. of big data—namely velocity and volume—pose challenges for educational organizations, especially in developing nations that do not possess the human and financial capacity to handle the velocity and volume of data. Sagiroglu and Sinanc argued that big data implementations need to be planned carefully and with an eye toward growth, as humans have generated more digital data since 2010 than ever existed in the thousands of years previous [19]. The authors cautioned that the size of the data can be confusing, and the technical expertise of staff can be limiting, leading to a sense of overwhelm paralysis. This results in a wealth of data collection but little analysis, and without any aim towards decision-making and practicable outcomes [19]. Additionally, Sagiroglu and Sinanc asserted that organizations must have the capacity to store data in the first place and the ability to organize that data in a way where multiple stakeholders can access and interpret the data accurately. In developing nations, there may not be the physical or cloud storage capacity to gather and analyze data in a timely manner, positioning these nations in a perpetual deficit state [19].

Nazarenko and Khronusova echoed many of Sagiroglu and Sinanc's concerns, suggesting that educational organizations must prepare years or decades in advance to support the type of data storage that is necessary for big data decision-making [15]. The authors also explained that without clear goals and educational outcomes, many under-resourced schools and educational organizations will struggle with understanding what data to gather, where to gather it from, how much to gather, and when data collection stops and data analysis starts [15]. Moreover, Nazarenko and Khronusova explained that it is increasingly common to be in a perpetual state of data collection without the expertise to analyze it. Educational organizations often experience difficulties when recruiting and retaining high-quality data analysts who are technically trained to analyze millions or billions of data points across many different data types and formats [15]. As a result, organizations may realize a sense of overwhelm paralysis with regard to the volume of data, the velocity of data, and the staff and planning to execute their goals. Additionally, without well-defined goals, many educational organizations may gather data that does not serve the mission or vision of the organization or does not substantially inform how educational leaders can improve the organization [15,19].

3.2. Permissions, Consent, and Privacy Concerns

Collecting and analyzing data is one element of using big data to make education decisions. Yet, before the data is gathered—especially at public entities and in countries with data protection laws—educational organizations must procure permissions and consent, while ensuring the private nature and confidentiality of most or all data. In this regard, cybersecurity and the safety of big data is paramount for educational organizations [14].

Regarding data storage, Wang asserted that "there is no shortage of concerns over how to store, process, and access student learning data while preventing those data from being abused or misused ... " while "student learning data are collected and stored in different silos—school district offices, online learning systems, and mobile devices—that are not connected to one another" [3] (p. 383). Additionally, Wang argued that "the growth of the Internet outpaces laws and regulations. To date, there has been a lack of Institutional

Review Board Protocols or federal regulations that protect human participants in large-scale social experiments on the Internet" [3] (p. 383). Here, educational organizations seeking to make decisions using big data are likely going to face privacy- and cybersecurity-related challenges [14], even if those challenges are mitigated at the beginning of the process because of the speed in which technology and the Internet advances.

Regarding student privacy, Nazarenko and Khronusova explained that "much information about student's behavior is classified like personal data that cannot be collected without special permissions. Moreover, tracking of student's activity needs to be expanded by their personal information, such as temperament type. However, many students are not interested in providing this kind of information" to their institution [15] (p. 678). Here, there are not only challenges with permissions and consent policies related to data, but there is no guarantee that individual stakeholders such as students will consent to have their data gathered, possibly straining relationships between students, teachers, and their educational organizations.

There is also the issue of how permission and consent and safety policies are communicated to stakeholders. Williamson argued that many members of educational organizations have no interest in or knowledge of big data, possibly confusing stakeholders regarding big data and its utility [14]. Similarly, Dishon explained that when education is so data-driven, both students and teachers may not know what data has been collected, by whom, and for what purpose [20]. This sense of confusion could deter educational stakeholders from engaging with big data policies and contribute to their uncertainty about what data is being collected and analyzed, possibly producing a feeling of surveillance which has been found to negatively impact teaching, learning, and a sense of belonging [3,14,20].

3.3. Data as a Dehumanizing Force in Education

Although big data inherently requires human input to exist, researchers have long criticized the fact that humans often use big data in dehumanizing ways, resulting in students, teachers, administrators, and other stakeholders feeling powerless and less autonomous in their education experiences. Nazarenko and Khronusova explained that at the post-secondary level, where class sizes may be larger, the lack of personal education and discussion between students and lecturers may be marginalized and replaced by an emphasis on big data to inform teaching strategies and practices, many of which may be automated and Internet-based [15]. Here, Nazarenko and Khronusova argued that students may unintentionally experience depression and a feeling of social isolation if their process and educational experience is too reliant on big data and too separated from human interaction with their teachers [15].

Dishon also argued that educational environments should be personalized to the point that data-driven decision-making does not infringe upon one's sense of a naturalistic learning environment [20]. However, as teachers and administrators continue to use data to make informed decisions, stakeholders may begin feeling as if they are numbers and not people, placing a wedge between a student and their teacher and eroding trust within this important relationship [20]. Similarly, Johnson reasoned that as educational organizations and individual teachers gather data to make decisions, students may feel that their privacy is violated to the point where they do not feel as if they are individual learners. Johnson continued by saying that big data can contribute to "relationships [that] can easily be seen as contributing to a collectivization of subject, where all are treated identically based on the assumption that they are all 'typical' students" [21] (p. 5), resulting in students feeling unnecessarily homogenized and unimportant.

Perhaps most importantly, big data and its ability to accomplish educational goals has been known to historically marginalize communities of color and those belonging to underrepresented groups. In this way, big data can be seen as a tool of educational inequity and not the other way around. In their discussion of big data and Australian education systems, Buchanan and McPherson described this phenomenon as the "datafication of the learner" [16] (p. 30). This datafication can weaken student-teacher and school-community relationships, thus marginalizing many stakeholders [16]. In a discussion of the critical use of big data toward racial equity, Gillborn et al. explained:

> Quantitative data is often used to shut down, silence, and belittle equity work. Whenever governments, employers, or educators are challenged on their poor performance in relation to an under-represented group, they will typically reach for statistics in an effort to show that they are really much better than you might think. [22] (pp. 174–175)

Here, the authors reason that many school systems' underserved students of color or other groups and the use of big data can be a mechanism of masking educational inequities instead of identifying equity gaps and stemming them [22]. Moreover, Gillborn et al. suggested that the way in which communities of color and other marginalized groups are not engaged with data collection and analysis further marginalizes these communities, placing the students in a position of being surveilled without being served [22].

3.4. Is Equity Possible?

As with any data-driven decision-making, the data itself can be flawed, and the deductions made from the data can be equally problematic. First, developing nations will likely encounter challenges gathering accurate data on their people, particularly as race, gender, and socioeconomic status is concerned. Many developing countries have social and religious systems that discriminate against people from certain racial and ethnic backgrounds [4,23], while other countries marginalize people from Queer backgrounds [24]. Here, many developing nations may not have leadership that values human beings equally or equitably, leading to large datasets that are incomplete or inauthentic according to someone's true, authentic identity.

When discussing how data can inform decision-making, Macfadyen et al. called for the need for a more effective overall assessment paradigm in education, as many data driven decisions are made using incomplete data and may inform targeted interventions that are not timely or efficient enough [18]. Similarly, Nazarenko and Khronusova asserted that some forms of data are much easier to collect than others, comparing electronic standardized test data to word-of-mouth communication [15]. The authors argued that word of mouth communication may be essential in understanding how educational organizations can implement change, but "verified data collection of this kind is practically impossible" [15] (p. 678).

Johnson backpedaled in their discussion of the integrity of big data, explaining that data mining and the problems with big data go deeper than poor methodology [21]. Johnson claimed that an inherent feature of science in technology is how data collection instruments and strategies are weaved into "a complex web of technical and social interdependencies," [21] (p. 7), such as administrator priorities, changing student demographics, and unsteady influxes of human and financial resources. Johnson, therefore, argued that "Design intent and assumptions about user behavior are especially significant sources of embedded values in technologies" [21] (p. 7), suggesting that educational leaders and policymakers must understand who implemented the data collection measures and which specific social forces may have influenced those approaches.

Regarding data driven decision making, Crossley raised important questions, including whether researchers "should ask whose capacity will be strengthened by new initiatives, whose values and approaches to research will be prioritized, whose modalities will be applied—and do these meet local needs, priorities and agendas?" [17] (p. 22). Using the European Union and the United Kingdom as an example, Crossley questioned the value of "expensive big science approaches to social research that are increasingly favored in the UK" and whether such approaches "have the best potential to foster the strengthening of research capacity within low-income countries" [17] (p. 22). Here, Crossley understood that what may be good for one educational organization or context may not be good for another, yet it may be tempting for developing nations and developed nations to overgeneralize data and its implications when individual data initiatives are best for a certain educational context [17].

Buchanan and McPherson elaborated on this false sense of data integrity when discussing Australia's national testing program to evaluate primary and secondary student progress and teaching effectiveness [16]. In their critique of Australia's national testing plan, the authors suggested that Australia had modeled their high stakes testing program against those from other developed nations, but such a strategy was not best for Australia, which is famous for its stark contrast between rural and urban school districts [16]. In all, Buchanan and McPherson argued that the primary justification for the testing program was to formalize some sort of mechanism that measures and produces good teaching, but the program ultimately equated "student achievement to a crude test result," [16] (p. 31), which did little to inform Australia's idiosyncratic school system at both the regional and national levels.

In U.S. contexts, Gillborn et al. criticized common uses of big data, again targeting national testing programs as Buchanan and McPherson did [22]. With a critical lens toward racial equity, Gillborn et al. argued that "National testing programs, such as the No Child Left Behind (NCLB) reforms in the US and the use of school performance tables in England, have popularized the idea that numbers can be used to expose (and change) failing schools" [22] (p. 161). However, as the authors reason, "commentaries [on these programs] rarely include any detail about the relatively small samples," [22] (p. 161), in some instances numbering only 200, yet the results were being generalized across tens of thousands of schools. In this regard, Gillborn et al. implied that data can be flawed or not measure what it is intended to measure, while the interpretation and implementation of that data to inform policy and practice can be equally harmful [22]. Furthermore, such neoliberal policies have redirected public resources to private sectors, impacting public educational funding and increasing the equity gap between low- and high-income communities, as well as exacerbating racial equity gaps in U.S. contexts [22].

3.5. Can Data Be Captured and Used at All?

As mentioned earlier, many educational organizations struggle to procure the necessary human and financial resources to capture and analyze big data. A plethora of research has asserted that one of the largest challenges facing educational organizations is recruiting and retaining high-quality staff to manage big data and perform data analytics [15,17]. In this case, educational organizations often compete with the private sector for personnel fluent in quantitative methods and machine learning. As Macfadyen et al. argued, "it may not be surprising, then, that globally, education lags behind all other sectors in harnessing the power of analytics," as "a preliminary analysis indicates that educational institutions simply lack the practical, technical and financial capacity to effectively gather, manage and mine big data," [18] (p. 22).

Moreover, as bureaucracies, educational organizations resist change and innovation, often embracing an organizational culture that clings to prior methods of operation [4,6,15,17]. Macfadyen et al. reasoned that "there is recognition that even where technological competence and data exist, simple presentation of the facts (the potential power of analytics), no matter how accurate and authoritative, may not be enough to overcome institutional resistance" [18] (p. 22). This resistance comes in several forms, namely a cultural resistance that is established and perpetuated by organizational leadership, yet resistance also comes in the form of a lack of human or financial capacity to change [15,17], and a resistance to embrace big data due to an inability to strategically plan goals and initiatives to use big data [5,14,16,22].

3.6. Turnover and Continuity

Big data collection and analysis may be computerized and automated by developed nations or individual wealthy schools, but for many developing nations, the business of big data is human intensive. This context requires consistent and highly skilled staff to gather and analyze data toward educational equity, with human beings able to understand nuanced educational inequities and paths toward remediating those inequities [14]. However, education has remained one of the fields with the highest turnover of personnel [10,18,25], meaning that the humanistic nature of big data collection and analysis is inherently inconsistent and continuously disrupted by teacher and administrator turnover at the primary, secondary, and post-secondary levels [14].

Macfadyen et al. outlined the unique problems facing the field of education because of leadership change, as educational organizations often adjust goals and strategic plans to align with new leadership, which implies that the collection and analysis of data is also likely to experience constant change toward new initiatives and future efficiencies [18]. Nguyen et al. echoed this sentiment, stating that teacher turnover often upsets data collection and analysis techniques, especially as teachers and administrators strive to meet accountability measures, whether at the regional or national level [25]. Often, individual schools must invest a considerable amount of human and financial capital to replace teachers, which then introduces new educational staff into a system and a potentially nuanced way of gathering individual student data [14].

Kraft et al.'s work suggested that teacher turnover was tied to student achievement data, also suggesting that as teachers leave the classroom, districts and school systems must track this teacher turnover and integrate this variable into big data sets to sufficiently control for this phenomenon and maintain the integrity of data-informed analyses [10]. Overall, research suggests that data collection and analysis is inherently humanistic, and teacher and administrator turnover at the primary, secondary, and postsecondary levels introduce changing variables (humans) into a complex system of big data, working against educational progress toward equity [5,7,25].

4. Uncertainties of Big Data in Education

Ultimately, considering the benefits and drawbacks from an era of big data in education that is already upon us, there are many uncertainties surrounding the future of the field. Perhaps most importantly, educational organizations must build the capacity to maintain pace with technology and the ever-looming threat of cybersecurity breaches and data loss. Unfortunately, government entities and educational organizations must work together to prepare for an uncertain future where one large data breach could threaten the very existence of a school and the data of countless stakeholders. Similarly, educational organizations should diligently capture data that does not require student or stakeholder consent, but these organizations should also develop a sense of trust between the organization and its community. Through establishing this trust, students and other stakeholders may be more interested in providing correct, robust data to allow the school or organization to make the best-informed decisions.

There is also the uncertainty of how big data can actually inform policy, or if big data will simply exist in the cloud or on a server without analysis, contextualization, and policy advocacy. Macfadyen et al. reasoned that, "the challenge of bringing about institution–wide change in such complex and anarchic adaptive systems may rightly be characterized as a 'wicked problem'– a problem that is complex, unpredictable, open ended, or intractable" [18] (p. 22). Here, the very nature of big data and the possibility of overwhelming paralysis or unsteady leadership could lead many big data initiatives down unclear pathways.

As of the writing of this review, many developing nations do not have big datasets nor the means to assemble them, and it is unclear whether developed nations will partner with developing nations to improve educational equity on a global scale. It often remains the responsibility of local or national governments to gather resources, learn from other nations, and launch big data initiatives. For example, in the Caribbean context, Charran et al. reasoned that literature on inclusive education in the Caribbean shows a deficit in the availability of special education services and resources and a lack of teacher training in special education [4]. Here, Charran et al. argued that the challenge for governments—

such as those of Caribbean nations—is creating education policies that detail specific educational interventions, such as providing appropriate special education services and mandatory teacher preparation for working with students with disabilities [4]. In this regard, government involvement is crucial, but without the necessary resources, developing nations may not be able to gather data, use it to identify interventions, and realize change, thus falling further behind developed nations [4].

Additionally, developing South Asian nations have also struggled with data collection and the public availability of data. Pakistan experienced considerable population growth in the 1980s and 1990s, resulting in consecutive decades of at least 4% population growth and comparable educational enrollment. However, Pakistan's Ministry of Federal Education and Professional Training does not make large or longitudinal education data available to the general public, and the Ministry's website does not house any large or longitudinal datasets at the primary, secondary, or postsecondary level [26]. As recent as 2013, Pakistani educational researchers have bemoaned the fact that Pakistani educational data is not available, claiming that, "To the best of our knowledge, there is none for Pakistan which uses a historical series of disaggregated data of education to investigate both level and growth effect of human capital on the economic growth" [27] (p. 384).

In the Middle East, ravaged by war and faltering national economies, Syria's educational system has been in crisis for over a decade, partially resulting from little or no regular data collection and analysis. In a Syrian educational report from 2022, researchers suggested that data collection, disaggregation, and analysis was a primary factor in limiting Syrian educational progress. As the researchers wrote, "Data in Syria is not disaggregated by hub or geographical region, demonstrating a challenge with data integrity across Syria, both within and beyond the education" [28] (p. 12). Even though Syrian educational leaders have attempted to gather data systematically in recent years in order to "facilitate better coordination across the different hubs, including in terms of information management, this has not yielded robust results in terms of data and data analysis within the education sector" [27] (p. 12). Here, many developing nations have endured years or decades of struggles in terms of meeting the basic needs of their citizens, never mind embarking upon educational data collection and analysis projects to make better informed educational decisions.

In all, educational organizations are operating in an increasingly complex and competitive environment where data is currency. Educational leaders are under increasing pressure to respond to shifts in national and local economies, as well as political and social change such as the growing need to increase access to education for low-income communities, communities of color, people with disabilities, and individuals from marginalized groups. Unfortunately, many developing nations serve large populations of marginalized people, and these are the very nations that could most benefit from big data initiatives to help educational organizations become more efficient and effective with fewer resources and more global competition for students and talent.

5. The Neoliberal Shift away from Educational Equity

In many nations, neoliberal policies and agendas have prioritized the growth of the private sector and the defunding of public goods, including educational services. Recent work has underscored the necessity for educational organizations to facilitate data collection and dataset construction initiatives [29], but these initiatives may prove futile if national or local-level governments are not supportive of educational initiatives or view privatization of education as preferable. Yet, a wealth of research has found that neoliberal education policies and practices often minoritize the neediest communities and students, including communities of color and communities from low-income backgrounds [30]. Although prevalent in developed nations such as the United States and members of the European Union, governments of developing nations have begun adopting neoliberal policies, leveraging data for supposed accountability purposes to justify a shift towards the privatization of educational services. Subsequently, socioeconomic and racial equity gaps have emerged in many educational settings, bringing into question the purpose of using data for educational decision making if the aims of those decisions are to dismantle education systems that serve the most underserved in the name of neoliberalism [30].

6. Conclusions

Whether educational organizations are ready or not, big data is already changing the global education landscape and increasing opportunities for those nations who can leverage big data to make data driven decisions. For many developing nations, the adage “you can’t manage what you don’t measure,” may ring true, while many impoverished school districts simply cannot measure what they cannot manage. School leaders and teachers are already under enormous pressure as it is, so asking these stakeholders to develop big data sets to inform the work they do seems particularly onerous. Additionally, many developing nations may be struggling with national-level concerns such as war and economic challenges that render educational data collection and analysis a potential afterthought [4,26–28,31].

As a result, developing nations should work alongside developed nations to build the human, financial, and technological capacity necessary to chart a pathway toward big data fluency and utility. Within developed nations, educational leaders are already enlarging big data and performing transnational analyses of big data to inform educational change on a global scale [5,7,9,11,25,31]. However, comparisons of developed and developing nations may prove futile, as comparing nations is not only difficult but perhaps nonsensical given the vastly different geopolitical and social divides between nations. Understanding these divides, developed nations also have a responsibility to perform the necessary equity work to partner with developing nations to ensure that this educational change is on a truly global scale and that is inclusive of all nations and their students, schools, and communities.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. SAS Institute. Big Data: What It Is and Why It Matters. 2022. Available online: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (accessed on 12 January 2023).
2. SAS Institute. Big Data: Who Uses It. 2022. Available online: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#industries (accessed on 12 January 2023).
3. Wang, Y. Big Opportunities and Big Concerns of Big Data in Education. *Techtrends* **2016**, *60*, 381–384. [CrossRef]
4. 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/Heidelberg, Germany, 2019; pp. 51–65.
5. Mandinach, E.B. A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educ. Psychol.* **2012**, *47*, 71–85. [CrossRef]
6. 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]
7. Schildkamp, K.; Poortman, C.; Luyten, H.; Ebbeler, J. Factors promoting and hindering data-based decision making in schools. *Sch. Eff. Sch. Improv.* **2017**, *28*, 242–258. [CrossRef]
8. McAfee, A.; Brynjolfsson, E. Big Data: The Management Revolution. Harvard Business Review. 2012. Available online: https://hbr.org/2012/10/big-data-the-management-revolution (accessed on 12 January 2023).
9. European Commission. European Education Area: Quality Education and Training for All. European Commission. 2022. Available online: https://education.ec.europa.eu/ (accessed on 12 January 2023).

10. Kraft, M.A.; Marinell, W.H.; Yee, D.S.-W. School Organizational Contexts, Teacher Turnover, and Student Achievement. *Am. Educ. Res. J.* **2016**, *53*, 1411–1449. [CrossRef]
11. Meschi, E.; Scervini, F. A new dataset on educational inequality. *Empir. Econ.* **2014**, *47*, 695–716. [CrossRef]
12. 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]
13. Schildkamp, K.; Lai, M.K.; Earl, L. (Eds.) *Data-Based Decision Making in Education: Challenges and Opportunities*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
14. Williamson, B. *Big Data in Education: The Digital Future of Learning, Policy and Practice*; Sage: New York, NY, USA, 2017.
15. Nazarenko, M.A.; Khronusova, T.A. Big data in modern higher education: Benefits and criticism. In Proceedings of the 2017 International Conference: Quality Management, Transport and Information Security, Information Technologies, Saint Petersburg, Russia, 24–30 September 2017. [CrossRef]
16. Buchanan, R.; McPherson, A. Teachers and learners in a time of big data. *J. Philos. Sch.* **2019**, *6*, 26–43. [CrossRef]
17. Crossley, M. Global league tables, big data and the international transfer of educational research modalities. *Comp. Educ.* **2014**, *50*, 15–26. [CrossRef]
18. Macfadyen, L.P.; Dawson, S.; Pardo, A.; Gašević, D. Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge. *Res. Pract. Assess.* **2014**, *9*, 17–28. Available online: http://www.rpajournal.com/dev/wp-content/uploads/2014/10/A2.pdf (accessed on 12 January 2023).
19. Sagiroglu, S.; Sinanc, D. Big data: A review. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013. [CrossRef]
20. Dishon, G. New data, old tensions: Big data, personalized learning, and the challenges of progressive education. *Theory Res. Educ.* **2017**, *15*, 272–289. [CrossRef]
21. Johnson, J.A. The Ethics of Big Data in Higher Education. *Int. Rev. Inf. Ethics* **2014**, *21*, 3–10. [CrossRef]
22. Gillborn, D.; Warmington, P.; Demack, S. QuantCrit: Education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race Ethn. Educ.* **2018**, *21*, 158–179. [CrossRef]
23. Telles, E. *Pigmentocracies: Ethnicity, Race, and Color in Latin America*; University of North Carolina Press: Chapel Hill, ND, USA, 2014.
24. 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]
25. Nguyen, T.D.; Pham, L.D.; Crouch, M.; Springer, M.G. The correlates of teacher turnover: An updated and expanded Meta-analysis of the literature. *Educ. Res. Rev.* **2020**, *31*, 100355. [CrossRef]
26. Pakistan Ministry of Federal Education and Professional Training. Pakistan's Ministry of Federal Education and Professional Training. Government of Pakistan. 2023. Available online: http://www.mofept.gov.pk/ (accessed on 17 January 2023).
27. Jalil, A.; Idrees, M. Modeling the impact of education on the economic growth: Evidence from aggregated and disaggregated time series data of Pakistan. *Econ. Model.* **2013**, *31*, 383–388. [CrossRef]
28. Qaddour, K.; Husain, S. Syria's Education Crisis: A Sustainable Approach after 11 Years of Conflict. Middle Eastern Institute. 2022. Available online: https://www.mei.edu/publications/syrias-education-crisis-sustainable-approach-after-11-years-conflict (accessed on 17 January 2023).
29. 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. [CrossRef]
30. St John, E.P.; Daun-Barnett, N.; Moronski-Chapman, K.M. *Public Policy and Higher Education: Strategies for Preparation, Access, and College Success*; Routledge: New York, NY, USA, 2018.
31. Taylor, Z.W. College Admissions for L2 Students: Comparing L1 and L2 Readability of Admissions Materials for U.S. Higher Education. *J. Coll. Access* **2020**, *5*, 54–67. Available online: https://scholarworks.wmich.edu/jca/vol5/iss1/6 (accessed on 21 April 2023).