Hidden Figures: Systemic Racism and the Possible Effects of Sampling Bias in Large-Scale Education Data Sets

ElizabethAnn S. Kelly

ABSTRACT

Significant influence over national educational systems and educational policies is exerted by large-scale international studies of education such as the Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Achievement (PISA). The impact of these influences on educational systems and student achievement is evidenced by the changes in patterns of cross-national and cross-cultural disparities in educational achievement over iterations of the test. This paper examines the import of national educational policymaker actions and potential sample bias of these tests through the lens of racism and educational inequities. Despite a data gap, national policy makers implement changes in educational policies and practices based on results from these large-scale studies and so fail to take into account the differing contexts students, especially minority cohorts, experience within education settings. Additionally, some specific examples of practices to exclude specific student cohorts from test participation are presented.

Keywords: Systemic racism, PISA, TIMSS, sampling bias, large-scale dataset

The Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Achievement (PISA) are large-scale international studies of education which for decades have exerted significant influence over national education systems and domestic educational policies around the world, creating

changes in educational policy and practice (Baird et al. 2016; Feniger 2020; Takayama 2008). The impact of these influences on educational systems and student achievement is evidenced by changes in patterns of cross-national and cross-cultural disparities found in large-scale studies of educational achievement (Halpern et al. 2007; Hedges & Nowell 1995; Hyde, Fennema, & Lamon 1990; Hyde, Lindberg, Linn, Ellis, & Williams 2008; Lindberg, Hyde, Petersen, & Linn 2010). However, while these studies and others like them have helped researchers better understand the contexts in which students are both successful and unsuccessful, there are still stories being left untold, and figures being hidden by intranational systemic racism and sampling bias.[1] Both TIMSS and PISA publish extensively on their sampling methods and possible biases. However, how many policy makers read this literature? In 2014, an international group of scholars and educators published an open letter to the director of the PISA assessment outlining issues related to national governments making education policy decisions based on PISA assessment results and an interest in climbing through the PISA rankings (https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics). This paper identifies why exclusion of minorities from the optimal benefits of education exacerbates the effects of racism in society and specific instances in PISA and TIMSS administration which reflect the institutional exclusion of specific minority cohorts which increase systemic inequity in educational assessment and the concomitant reforms emanating from national results.

THE VALUE OF EDUCATION

Learning is valued cross-culturally by policy makers, and education is considered a basic human right by the United Nations: equal access to education is the fourth goal in the United Nations' plan for a sustainable future (United Nations 2019). Despite this, and much intervention on the ground, rates of attendance and graduation from secondary education have not risen in a num-

ber of low- and middle-income countries contained in UNESCO data, as well as in African focused cross-national datasets (UNESCO 2017; Results report 2019, Global Partnership for Education results report). An important critique of current policy initiatives in education is the tendency to view education as a "cure-all" for inequities, without first addressing the inequities inherent in the educational system (Datzberger 2018). We cannot approach education as isolated from the political, social, and economic context in which it occurs, a context that embodies systemic racism emanating from power dynamics as well as institutional "efficiencies."

To better understand the structural changes required, it is necessary to conduct research investigating the contexts in which students are being educated. The seeds of the disparity in educational achievement may be found in Michel Foucault's exhaustive writings on the concept of power, where he posited that power comes from knowledge, uses knowledge, and then power reproduces itself through the creation of further knowledge. Power and knowledge cannot exist in isolation, but only in mutuality. To that end, he coined the neologism "power-knowledge" (pouvoir-savoir) (Foucault 1976).

In more recent research, this idea of power-knowledge within individuals is conceptualized as human capital, which is defined as the total knowledge, skills, and experience that a person possesses (Aslam 2014). Higher human capital is associated with a sense of empowerment, of which there are several kinds, including economic, social, psychological, and political. In many cases, equality largely means reducing the constraints on the lives of minority individuals, while simultaneously increasing the number and types of opportunity offered to them, resulting in economic and political empowerment, the mechanism through which education influences outcomes. Therefore, in order to affect real change in society through education, and reduce inequities, it is necessary to start formulating education policy from a transformative rather than an assimilative perspective (Datzberger 2018). Most current educational policy is written from an assimilative

perspective: policy makers believe that by giving children equal opportunities for education, they will then have equal opportunity in life (Sachs 2015). However, this is an overly simplistic view of educational reform. Instead, a transformative approach to education is needed, one that provides equity, not merely equality, in opportunities for education (Young 2001).

POWER THROUGH EDUCATION

An education is required to achieve economic empowerment in most segments of the globe, as that education provides the entry into higher paying careers (Aslam 2014). As level of education increases, a person's income also often increases. This association is stronger for women than men who are afforded the same educational advantage. In Bangladesh and Pakistan, this increase is double for women what it is for men (Aslam 2014).

Political empowerment is the extent to which a person can effect real change in their society and government. There is a long-standing power imbalance between the educated and the uneducated in politics; it is even codified in the 12th Amendment of the United States Constitution, which describes the way in which the President and Vice-president are elected via the Electoral College in lieu of a plebiscite. Hamilton's commentary on the role of the Electoral College in the Presidential election makes clear that the express purpose is to ensure the disenfranchisement of the uneducated poor (Hamilton 1788). Interestingly, while this discrepancy in political power between the educated and the uneducated persists, it is lessening due to the rise of social media and the Internet. Through the use of online platforms, people with less education are more likely to perceive an increase in political empowerment (Sasaki 2017).

EDUCATION AS A TOOL

The political power imbalance between the educated and the uneducated is one of the underpinnings of educational psychology as a field (Dewey 2008). Accepting Foucault's (1976) concept of power-knowledge impels the conclusion that education (in

that transfer of knowledge is the mechanism by which education achieves its goals) is an analog of the transfer of power in our society; thus it is fraught in its potential to consolidate or shift control away for those currently exercising control over society. While a tool of empowerment, educational access and content historically has been manipulated in more malignant ways, too (Blackburn 1987; Ingram 2013; Lewis 2017; Rostam-kolayi 2014; Watts 2013).

Education as Oppression

Cross-nationally, education at times has been used as a tool of oppression, in limiting the free development of students to become their full and best selves, in targeting their behavior and restricting their actions, as well as in constraining their spheres of interest and studies (Blackburn 1987; Ingram 2013; Lewis 2017; Rostam-kolayi 2014; Watts 2013). The education of girls and minorities often has been conceived to restrict them, rather than to create active participants in the world (Blackburn 1987; Connell Szasz 1980; Maina 2006; Watts 2013). While character development and the engendering of virtue in students has been a historical goal of mass education globally (Benavot & Resnik 2006), there has been specific, and limiting, emphasis on what was acceptable for specific groups. Parity in enrollment rates is not sufficient to achieve equity for less powerful cohorts; studies show that even when enrolled in school girls, rural students, and those from racial and social minorities do not receive the same level of learning as majority boys (Hickey & Hossain 2019; World Bank 2017). For example, a randomized controlled trial of teacher bias in mathematics found that while teachers do not differ when correcting work, when evaluating student mathematics ability they assess students with stereotypically female and nonwhite names as having lower abilities than stereotypically white and male names, even when the actual achievement scores were the same (Copur-Gencturk et al. 2020). Additionally, while membership in a minority racial or ethnic cohort does not necessarily correlate with lower socioeconomic status (SES) across

all global populations, when those students do fall into a lower SES category, they along with the rest of the cohort are likely to be treated differently: a study in Australia found that lower-SES schools were not only less likely to offer high-level mathematics subjects, when they were offered students were far less likely to choose those courses than their peers living in higher-SES neighborhoods (Murphy 2019). Those who are Black and female, at least in the United States, may undergo an even more disadvantaged educational environment.

Black Girls in the United States

Being black and female creates a lived experience that differs from peers in the United States. In schools with a majority of minority students, Black girls who perform well academically are often criticized for their behavior (Morris 2007). Teachers target Black girls, focusing on improving their etiquette and encouraging them to ascribe to ideals of white femininity, showing the impact of the intersection of race, class, and gender on girls in school (Morris 2007). These experiences in the classroom shape how girls perceive themselves and their capabilities; even the most academically gifted girls feel inferior to boys in mathematics and science when teachers focus on behavior and signal more appreciation of boys' effort (Spearman & Watt 2013). These perceptions of inferiority can be found in girls as young as six. In first grade girls often begin avoiding educational activities that are described as "for smart kids"; this behavior is not found in boys (Bian et al. 2017).

Education as Colonization

While it cannot be denied that religious institutions' interest in the morality of girls helped to formalize and expand girls' education in Europe and the Great Britain, there is a darker side to religious institutional involvement, as religious-based educational systems also were used as a tool of colonialization outside of Europe (Churchill 2004; Mohanty et al. 1991; Mujuni 2015; Turyasiimwa 2020), with specific and separate effects on female students.

Uganda. Prior to British colonialization in 1877, what is now the country of Uganda had no formal education system; educational standards and expectations were set locally within each formal or informal governmental unit (smaller kingdoms and tribal lands) (Mujuni 2015). One of the leaders of the region, concerned with the threat of colonialization, invited missionaries to implement formal education in 1875 as means to resist potential invasion (Mujuni 2015).

Due in large part to the inroads made by the missionaries, Uganda became a colony of the British Empire in 1894, less than 20 years later (Turyasiimwa 2020). Colonization of Uganda resulted in a formal state educational system for boys, but no provision was made under British rule for the education of girls, due to the local culture's views on the superiority of men and the influences of the missionaries, which continued through the 1960s (Sekamwa 2000; Turyasiimwa 2020).

Residential Schools. While formal education created a pathway for colonization in Uganda, it was explicitly used as a tool of colonization and genocide in Canada and the United States, or in the words of the creator of the residential school system in the United States it was "education for extinction." The stated goal was to eradicate all vestiges of indigenous cultures in North America (Churchill 2004). To that end children were forcibly removed from their parents and placed in residential schools where they would be educated to be "white" and punished for practices that were deemed "heathen."

Much of the focus of research in this area, like much history of formal education as a whole, has focused on boys; however, one of the more concentrated initiatives undertaken by the American Bureau of Indian Affairs was the removal of girls from their homes to "re-educate" them into "ideal women" (Trennert 1982). By focusing on cutting off girls from their cultures these men were effectively interfering with the generational transmission of cultural practices and knowledge from mother to child, which is particularly insidious when considered in the context of the

matriarchal structure of many North American Indigenous cultures (Churchill 2004). This also meets the United Nations definition of genocide by both removing children from one group to another, and by inflicting "conditions of life" meant to destroy an existing culture (Convention on the Prevention and Punishment of the Crime of Genocide, 1948). In addition to the genocidal intent of the men responsible, increased military regimentation, and lack of recognition of the reservation environment to which the children would return, contributed to the failure of the schools to meet the educational needs of Native American students (Trennert 1982).

To further their "education" many girls at residential schools were placed in the "outing" system established to introduce Native American young women to the running of a home, with all its attendant chores; this devolved into servitude in white homes with the proceeds used to support the costs of running of the residential school and providing no real benefit to the girls, forcing them into slave labor (Paxton 2006). Additionally, the girls labored to maintain the domestic environment at the schools themselves, to further reduce costs (Trennert 1982).

While the conditions within the United States Residential School System were truly deplorable, they pale in comparison to the conditions found in Canada. In concert with various Christian institutions,[2] the Canadian government oversaw the forcible removal and education of over 150,000 indigenous children from 1920 until the last residential school closed in 1996 (MacDonald 2015). Similar to the American system, however, no balance existed between manual labor to keep the schools running and the ostensible education for which the schools were established. Rather than providing an education to benefit their charges, let alone meet equivalency with their white counterparts, estimates indicate in the early decades of operation less than 50 percent of the boys and girls even survived, often due to diseases such as tuberculosis, or persistent malnutrition and its attendant health impacts (Milloy & McCallum 2017). The Canadian government

not only engaged in the cultural aspects of genocide, but in failing to protect the health of the children they kidnapped from their families, the government is responsible for the systematic extermination of indigenous children. Similar seeds of institutional racism have grown runners which run through today's educational systems.

RACISM IN EDUCATION TODAY

Historically, education has been wielded by the powerful to achieve their goals, to the benefit of some and detriment of others. Access to education has a potentially transformative effect for minority cohorts, allowing increases in social and economic empowerment. Race and minority ethnic status could be an important predictor in educational achievement in international testing, however this data is not gathered by PISA or TIMSS administrators for most participant nations.[3] Sociologists and other researchers have documented policies, processes, cultures, attitudes, and events that have led across the globe to racist outcomes, including exploitation, violence, discrimination, and intolerance. The complexity of race and racism, as well as its subtleties, have been investigated and interpreted across cultures (Andrews et al. 2014), from attitudes towards blood, skin color, and sport in the United Kingdom (Yuval-davis et al. 2009) or immigrant youth and sexuality in Sweden (Bredström 2003) or nation-building in both post-colonial Southeast Asia and east Asia (Ang 2022). Racism goes beyond color, especially the overly simplistic bifurcation of white people or people of color, as white-on-white racism has been documented repeatedly (an example of this would be the institutionalized racism Romani and Irish Travelers face in the UK and Ireland). Researchers found an increasing impact of racism in middle tier countries (those not falling within the bounds of traditional Western countries, or the poorer, Western-colonized countries) which exploit poorer countries or their own ethnic minorities (Dunaway 2016). Despite this extensive scholarship, neither TIMSS or PISA collects data on

race. The inclusion of race in the data collected could provide the ability for more nuanced explorations of the dataset and the basis for more equitable educational policy changes. Related to race, while Primary Language Spoken at Home is collected for TIMSS, no data is collected on if students are first- or second-generation immigrants, which has been shown to be predictive of achievement in individual countries.

LARGE-SCALE EDUCATION DATASETS

In an effort to understand the "whats and hows" of education around the world, large-scale studies such as TIMSS and PISA have been developed and have become a central part of the study of comparative education. In addition to providing snapshots of the practices and contexts of education, these studies can create impetus for change at the national level when policymakers examine national results, especially in relationship to results of other countries.

TIMSS

TIMSS data are gathered under the auspices of International Association for the Evaluation of Educational Achievement (IEA) which designed and conducted education-related large-scale comparative studies over the past six decades. Intended to assess changes in student performance related to science and mathematics, TIMSS has measured student achievement and collected comprehensive contextual information from teachers, students, and principals at four-year intervals over the past twenty-five years, being first administered in 1995 (Mullis et al. 1997); it was administered most recently in 2019 (Mullis & Martin 2017). The purpose of the TIMSS is twofold: first, it aims to assess global trends in STEM instruction; and second, it informs educational policy at both the national and international level (Mullis et al. 2016).

TIMSS's student performance data can be disaggregated in myriad ways; because it is accompanied by significant student, class, and institution background information, specific conditions affecting student achievement can be determined (Broer et

al. 2019). A two-stage random sample design is utilized for participation in a TIMSS cycle, employing representative and well-documented probability samples (LaRoche et al. 2016b).

Like other large-scale assessments, TIMSS has evolved over its existence. The instrument was developed based on methods initially used by the National Assessment of Educational Progress (NAEP)s, and focuses on curriculum-based achievement rather than knowledge-based achievement. In part because TIMSS is international, cultural and educational system differences required adaptations to the design, analysis, and subsequent reporting of findings from the assessment (Martin & Mullis 2019). Much has remained the same through subsequent TIMSS cycles, but the survey has evolved as variations in sample design, survey administration, and national participation have occurred. For example, in early surveys multiple adjacent grades were assessed in order to target a specific age (13), but in subsequent surveys only eighth graders were included (Broer et al. 2019). Recent editions included parental questionnaires for fourth graders (Broer et al. 2019). The instrument also has changed regarding inclusion of specific questions when their validity has been found suspect (e.g., elimination of parental occupation and income) based upon issues found in other IEA studies (Buchmann 2002). The questions now also recognize the impact of developments in technology, such as increased internet accessibility or computers in the home (Broer et al. 2019). Finally, different national educational systems elect to take part in each cycle of TIMSS. Research shows that participation in various IEA large-scale assessments is related to previous participation in similar assessments and overall wealth of the country (OECD, 2015).

PISA

Similar to the IEA, the Organisation for Economic Co-operation and Development (OECD) designed the PISA as a large-scale, cross-national survey, but instead focuses on measuring student knowledge rather than grade level curriculum. The goal of PISA is to assist governments in evaluating students' ability to apply

acquired knowledge based on literacy testing for reading, mathematics, and science, as well as problem-solving skills, independent of the national educational system curricula. Beginning in 2000, PISA has operated triennially, with each of the three subjects being tested each cycle and one being of particular focus in a given year (reading in 2000 and 2009, mathematics in 2003 and 2012, and science in 2006 and 2015.) The 2018 iteration added a new dimension to assess digital literacy, including distinguishing between fact and opinion (Schleicher 2019). PISA's two-stage sampling process first identifies a diverse cohort of schools estimated to be representative based on location and demographic factors (e.g., rural or urban) then randomly selects 40+ students to sit for the exam. A range of four to eight thousand students are surveyed for each country, and each student is assigned a sampling weight to reflect the nation's PISA-eligible class (Schleicher 2019).

When the PISA instrument is compared to TIMSS, the students are on average older (15-year-olds vs. eighth graders), and expected skills are tested instead of assessing achievement based on grade level curricula. Additionally, the tests use dissimilar scaling techniques based on different models of item response theory, as well as differing in test length and focus of student questionnaires (He et al. 2018). There are some significant overlaps in the measured constructs of TIMSS and PISA, however. For example, in both the 2015 PISA and TIMSS surveys, items on the background questionnaires provided to principals, teachers, and students contained substantially similar wording related to the same theoretical concepts (e.g., to assess the context of the learning environment, both instruments used Likert scale items on motivation, subject matter enjoyment, and engagement with the school community) (He 2018; OECD 2015; Mullis 2013).

The use of PISA data to assess achievement at the national level and create educational policy changes and curriculum development has been well documented (Breakspear 2012; Meyer, Heinz-Dieter, & Benavot 2014; Sellar 2014), even generating the term "PISA shock" when countries find the ranking of

their students' performance lower than expected (Elliott 2019). A number of issues arise from generalizing the success of specific educational practices from one nation or region to another, including some specific to PISA (Auld & Morris 2016; Hopfenbeck et al. 2018). External systems and practices may influence student achievements that are not captured in PISA datasets and that confound cross-national achievement comparisons (Feniger and Lefstein 2014; Alexander 2010). For example, the PISA background questionnaire does not ask students about outside tutoring, which is prevalent in Asian nations with a strong emphasis on national examination preparation and where a culture of extra tutoring outside of the classroom has evolved (Gillis 2016). This unobservable data could significantly raise performance on PISA, which tests for knowledge the student is expected to have, rather than assessing knowledge specifically tied to grade level curriculum, like TIMSS. Additionally, with PISA's sampling technique (selecting a small cohort of students from each school), and not tying teacher questionnaires to the specific learning environment for the tested student, the actual classroom inputs and teacher traits cannot be assessed, making it problematic to serve as a contextual predictor of achievement and the basis for implementing educational practices. Despite this, OECD makes teacher level recommendations based on PISA data (OECD 2010). Carnoy documents an issue regarding judging the effectiveness of educational practices by using TIMSS and PISA assessments for the same students one year apart, finding the predictive benefits of teacher and classroom characteristics (e.g., teacher quality and opportunity to learn) related to achievement to be overstated, at least at the country level (Carnoy 2016).

SAMPLING TECHNIQUES IN TIMSS AND PISA

Both TIMSS and PISA use a two-stage sampling strategy. First a random list of eligible schools is drawn up for each country, then for TIMSS random selection of one or more classes of the appropriate academic level are selected from each school. Sampling is

done by class rather than by student to allow for instructional environments to be used as variables (LaRoche et al. 2016). PISA samples students, not classes, therefore the second stage of PISA's sampling is selecting a cohort of students from the target population within the selected school. These students may or may not share classes or grade levels (OECD 2019).

For each cycle of TIMSS two datasets are collected, for fourth grade and eighth grade (or the country-appropriate equivalents); these represent the midpoint of primary education and the end point of primary education. In 2015, educational systems could choose to participate in either the fourth-grade assessment, the eighth-grade assessment, or both (LaRoche et al. 2016). Each participating national education system undertakes the test administration and collection of data, governed by documentation and training from the international project teams. In order to prevent bias in responses which could affect national outcomes, the IEA applies participation or response rate standards to participating education systems' data. Issues with response rates at the school, classroom, and student levels may lead to the exclusion of that system's data in the TIMSS and TIMSS Advanced international database and resulting reports (LaRoche et al. 2016). However, it is likely that some countries also use the TIMSS and PISA sample exclusion criteria to artificially inflate their TIMSS scores.

Both TIMSS and PISA designers have established benchmarks and methodology to ensure sample validity ensuring statistically defensible results. However, in addition to ignoring the illustrative and potentially transformative data on race, countries systematically exclude specific cohorts, ostensibly for efficiency or other administrative purposes. Such exclusions suggest the potential for sample bias, and additional harm when national policies might be promulgated subsequently which affect the excluded group. An in-depth analysis of each country's sampling for TIMSS and PISA is beyond the scope of this article; however, some examples are illustrative of the issues.

While research indicates student achievement on international assessments can be related to characteristics of students or

their classroom environments (Kelly 2022; Meinck 2017; Grace & Thompson 2003; Mullis 2012b), certain types of students are consistently excluded from participation in TIMSS and PISA testing. When IEA-established participation standards are contravened, this results in annotation or segregation of TIMSS reported data in order to highlight the potential validity issues (Mullis 2012a, Meinck 2017) and the OECD survey design includes similar sample validity standards and practices. Clearly, impossibility of test administration in specific circumstances meets a reasonability standard, however, national practices related to these excluded cohorts may exceed that standard.

Special Educational Needs

Both TIMSS and PISA allow for exclusion of schools dedicated to educating those with disabilities or specific students with disabilities. However, by allowing exclusion of students receiving their education in specialized environments or having a learning disability, even if the individual student is performing at "grade level," TIMSS and PISA are reinforcing the idea that special education is "less than" standard education classrooms. This emphasizes a culture of exclusion applying to special needs children, indicating that they are not participants in the cohort of achievement (Schuelka 2012). While this does not fall within the bounds of systemic racism, except to the extent race influences students' characterization as special needs, it does speak to systemic inequity. Symbolic annihilation applies to the erasure of segments of the population in popular media (Gerbner 1972), and, in a sense, students participating in special education programs or possessing a physical or educational disability similarly can be siloed and excluded from these international assessments, without any documentation that the particular disability would prevent valid participation. In essence, they become invisible.

Eliminating specific cohorts from TIMSS and PISA testing presents potential bias considerations and implications for international educational policy (Schuelka 2012, McGrew et al. 1992). Countries' use of discretion related to special needs stu-

dent participation results in a lack of transparency and therefore accountability (Gamazo et al., 2019) and countries' focus on high achievement for their students could exacerbate the inclusion of potentially low-performing students in special education settings which do not participate in the assessments. In the U.S., students with certain background characteristics have been funneled into special education programs, without regard to actual disability, in order to raise test scores (Reschly 1993; Vanchu-Orosco 2012). The potential is there for countries to behave similarly in order to influence scores on TIMSS, PISA, and their own national assessment surveys.

In the administration of PISA assessments, OECD allows for the participation of students whose disabilities are not severe enough to support exclusion; this can be accomplished through use of the standard test instrument or a revised one designed for those with learning differences. However, in 2015 only 11 countries employed that option (Gamazo et al., 2019). Additionally, analysis of PISA data finds a lower participation of special education students than occurs in the population, and, for some countries, increased individual student exclusion rates. Canada, in particular, while having a low school exclusion rate (for schools established specifically to meet the needs of students with disabilities), in 2015 excluded 7.1% of individual students in participating schools. Luxembourg and Norway follow a similar pattern of rising percentages of student exclusion (Brzyska 2018). TIMSS 2019 data reporting did not segregate excluded students with disabilities from those students speaking an alternative language; however, 17 participating countries had student exclusion rates at either the fourth grade or eighth grade level that exceeded the 5% benchmark, with some exceeding 10% (LaRoche et al. 2019). Despite exclusion from test participation, students with disabilities may be subject to national policies promulgated based on results from TIMSS and PISA assessment, which may not be optimal for their achievement. It should be argued, therefore that it would be more beneficial to include these students and allow researchers to examine their results in order to inform better policy initiatives.

Georgia

In addition to excluding students with special educational needs, countries have the discretion to only offer testing to students educated in a specific language, even when significant numbers of students are educated in another language. While not anti-BIPOC racism, language limitations in test administration may represent a facet of educational discrimination against ethnic minorities. In the case of Georgia, only including native (and thus ethnic) Georgian speakers in the target populations in both the 2015 and 2019 iterations of TIMSS, as well as the 2018 PISA assessment, results in a sample that is only representative of 90% of the target population (Mullis et al. 2017, UNICEF 2019). Ethnic tensions in Georgia have been documented extensively in the South Caucasus region and extend back decades; some analysts indicate they began in the 1990s, others say as far back as the 1920s or even earlier (Tabatadze 2016; Nichol 2009). While there is debate over Russia's actions in the past decades and the effects on intrastate conflicts, fundamentally there are ethnic factions of the Georgian state living in disharmony: independence-desiring Georgians and those in regions which identify or claim kinship with the Russian Federation, such as Abkhazia or South Ossetia (which are currently Russian-occupied) or are part of minority communities (Isakhanyanm 2012, Nichol 2009). Ethnic minorities constitute 16% of the Georgian population, with Azeris representing 6.3% and Armenians 4.5% (Georgian Census 2014). A rise of nationalism in the last part of the twentieth century exacerbated ill-treatment of minority populations, with ethnic Georgians maintaining rights and privileges that were not enjoyed by other residents (Wheatley 2006).

After becoming an independent nation, and after its subsequent admission in April 1999 to the Council of Europe, Georgia committed to ratification of both the Framework Convention for the Protection of National Minorities (FCNM) and the European Charter for Regional and Minority Languages. After long delays, the Georgian Parliament finally approved the FCNM and it became operable in April 2006. Despite adopting these

principles, initially the prevailing attitude by the majority government upon achieving its independence from Russia was that minority populations should react with gratitude in being allow to stay (Wheatley 2006).

The Georgian government has acknowledged its difficulties in achieving full integration of its diverse ethnic populations (UN Document CERD/C/GEO/4-5 2011), including in the area of education. The educational barriers are primarily inherent and systemic, relating to factors not in students' control (Kitiashvili 2016.) In 2017, in response to the United Nations report of its deficiencies in addressing the needs of minority populations, Georgian representatives noted they had implemented ethnic minority access to education at all levels in the students' native language; as of 2017, there were 220 non-Georgian language schools (Georgia, Addendum 2017). Providing native language education offers both benefits and negative outcomes for Georgian students. Research has found that while there is improvement in access to higher education for members of ethnic minority cohorts, limitations continue to restrict participation of these students. Additionally, by not including non-ethnic Georgians in the TIMSS or PISA samples, it is largely impossible to assess if these students are receiving a fair and equal education, as the Georgian government is claiming to provide.

Canada

A possibly more subtle introduction of sample bias occurs in Canada. According to the TIMSS International Results in Mathematics for both the 2015 and 2019 iterations, the Canadian sample is only indicative of 67% and 79% of the targeted population respectively. In Canada only two provinces participated in the 2015 eighth grade TIMSS samples, Ontario and Quebec; in 2019 students from the provinces of Alberta, Manitoba, Newfoundland, Ontario, and Quebec participated in the fourth grade assessment. While they are the most populous provinces, they are also the most affluent and whitest.

The 2018 PISA assessment specifically excludes Indigenous students living on reservations in Canada. Canada currently has over six hundred recognized First Nations governments and indigenous people make up the majority of the population in Nunavut and parts of the Northwest Territories. However, the tested provinces only have 2% to 3% indigenous populations. As an indigenous population, the First Peoples of Canada have and continue to experience exploitive and inequitable treatment and outcomes, in comparison to those of white ancestry across multiple spheres including education, environmental justice, food security, criminal justice, and heath care (Brzozowski et al. 2006; Deaton et al. 2020; Frost 2019; Hammond et al. 2017; M. Hu et al. 2019; Smylie et al. 2010). Given the Canadian government's history in the "education" of indigenous children, including the forcible removal of children to residential schools into the 1990s, this lack of inclusion of indigenous majority areas of Canada is deeply concerning, as well as likely masking the actual variability in achievement within the Canadian sample.

China

Similar to Canada, in China only certain provinces participate in PISA assessments; in 2018 these provinces were Beijing, Shanghai, Jiangsu, and Zhejiang (OECD 2019), which are four of the six wealthiest provinces (IMF 2022). China uses a hukou system, which defines if a person is from an urban or rural area based on their parentage. While it is now possible for rural hukou holders to work in cities, people must access certain government services through their hukou. As a result, rural hukou holders cannot access those services while working or living in cities (Brugiavini et al. 2018). This system has created significant differences between rural and urban individuals in terms of access to healthcare, education, and pensions or other benefits (Hung 2022). Research has shown that local residents are primed from childhood to be less trusting of non-local residents, especially in urban areas, resulting in a barrier for social integration for rural hukou

holders in cities (Luo & Wang 2020). These barriers increase for migrant ethnic minority individuals (Tan et al. 2022).

There are two issues related to the hukou system in the PISA sample; the first is the urban-rural divide for individuals who remain in their localities. By sampling only highly urban and wealthy areas, children from poorer rural areas are excluded from the PISA assessment. From 2010 to 2012, 100% of urban children completed primary and secondary education, while only 88% and 70% of rural children completed primary and secondary education, respectively. For that same time period, only 2% of rural individuals accessed any tertiary education, while 54% of urban individuals were able to receive tertiary education (Zhang et al. 2015). Such large gaps in educational attainment suggest that there are similarly large gaps in educational achievement between rural and urban children.

The second issue is that children with rural hukous who are living in cities cannot access secondary education without returning to their assigned localities. Subsequent to significant reforms to the hukou system regarding education, children belonging to migrant families may attend primary school where their parents are living and working, which has increased access to primary education for rural hukou children, but in order to access secondary education these children must return to the area from which their parents come (Zhou & Cheung 2017). Officially, these rural children still have access to the same secondary and tertiary educational opportunities as their urban peers, they must simply access those opportunities in another area (Hu & West 2014). This means for the purposes of PISA these rural children are not members of the target population, as they are excluded from the educational systems in the provinces being sampled, even if they are living the area.

CONCLUSION

It is widely accepted that access to education is key to individual empowerment, and most advocates also recognize that sys-

temic inequities in educational systems prevent minority cohorts from obtaining optimal benefits from these institutions. Recent backlash against critical race theory and inclusive educational practices in certain developed countries suggests that while this disadvantage is recognized, it is seen as a benefit by some. To change the status quo and create a more equitable global community, better data that is not "colorblind" is optimal to inform policymakers' decision-making processes, and also remove plausible deniability.

Both the TIMSS and PISA have significant influence, and they are rightly the benchmark for high-quality data collection at such a large scale. However, there has been significant overreach by policymakers in interpreting the actual implications of testing results regarding successful contexts for student achievement. While these assessments do provide valuable insights into global educational practices, allowing participating countries such broad oversight in sampling procedures has led some in practice to cherry-pick student populations in order to artificially enhance national educational profiles. Even in countries where explicit cherry-picking is not occurring, there has been very little done to address potential issues regarding systemic racism in educational systems and its influence on sampling procedures.

A tool in addressing systemic racism and other institutional inequities would be elimination of practices excluding minority cohorts from testing participation. Such changes would have to include sampling in rural areas that are more logistically complex to access, such as described in Canada and China. Additionally, test process modifications would require sampling students in special education environments, as students of color are disproportionately assigned to those classrooms. Progress also would necessitate that, in addition to these sources of systemic racism in education being recognized, they be addressed; that educational systems be continually studied to identify potential sources of systemic disadvantage so that they may be addressed both within research, and in the real world.

In reviewing the impact of these large-scale international achievement tests, other modifications of testing procedures are indicated to provide additional datapoints to understand the impact of racism in education. The introduction of collecting of racial demographic information could be a valuable way for PISA and TIMSS to inform public policy, and lead to increased transparency and accountability for educators and institutions. In contrast with other nations, the United States PISA sampling does gather data on race and ethnicity, which is a key tool in assessing racial achievement gaps within and between state education systems before students reach the age to take SATs. Collecting this information on a global scale would add an important missing context to the datasets as a whole. It is common for more homogenous countries to believe that racism is not an issue in their society. However, as discussed above, research indicates this is not accurate. By gathering racial and ethnicity characteristics, the information necessary to analyze operational aspects and outcomes by researchers could better inform public policy in education and help to hold institutions responsible for systemic inequities to ameliorate those conditions.

The current sampling systems employed by PISA and TIMSS allow countries to hide the statistical realities in some areas and so, too, the child figures behind them. As educators and researchers, we have a duty of care to address these sources of systemic inequities in our societies.

ENDNOTES

1 The intent is not to imply statistical invalidity of these tests, but that the methodology which permissibly does not include certain segments of the populations may result in sample bias which then presents issues when educational policy makers interpret results and generate implications beyond the data's actual explication.

2 In July 2022, Pope Francis visited Canada with the purpose of apologizing for the Roman Catholic Church's role in the genocide of First Nations Peoples within Canada. The visit had mixed reactions, with many members of First Nations stating the Pope did not go far enough

by not recognizing the Church's culpability as an institution, instead focusing on the actions of "individual 'Christians'"(Paradis 2022). Additional criticism was drawn by the Pope's refusal to rescind the Church's Doctrine of Discovery (Palmater 2022), which states Christians have the right and obligation to claim any and all lands not already inhabited by Christians (Romanus Pontifex 1455).

3 The United States PISA sample does include data on the race of participants, however this is not standard across national samples.

REFERENCES

Andrews, K., Bassel, L., & Winter, A. 2014. The British Sociological Association Race and Ethnicity Study Group Conference "Mapping the Field: Contemporary Theories of Race, Racism and Ethnicity." *Ethnic and Racial Studies* 37(10): 1862–1868. https://doi.org/10.108%1419870.2014.931994

Ang, I. 2022. "On the Perils of Racialized Chineseness: Race, Nation and Entangled Racisms in China and Southeast Asia." *Ethnic and Racial Studies* 45(4): 757–777. https://doi.org/10.108%1419870.2021.1990979

Aslam, M. 2014. *Empowering Women: Education and the Pathways of Change; Background paper prepared for the Education for all global monitoring report 2013/4, Teaching and learning: achieving quality for all; 2013*. http://www.worldbank.org/mdgs/gender.html

Benavot, A., & Resnik, J. 2006. "Lessons from the Past: A Comparative Socio-Historical Analysis of Primary and Secondary Education." *Global Educational Expansion Historical Legacies and Political Obstacles*, eds. Aaron Benavot, Julia Resnik, Javier Corrales (Cambridge, MA: American Academy of Arts and Sciences), 1–90.

Bian, L., Leslie, S. J., & Cimpian, A. 2017. "Gender Stereotypes about Intellectual Ability Emerge Early and Influence Children's Interests." *Science* 355(6323): 389–391. https://doi.org/10.1126/science.aah6524

Blackburn, J. 1987. "Schooling Women." *Australian Feminist Studies* 2(4): 213–219. https://doi.org/10.108%8164649.1987.9961554

Bredström, A. 2003. "Gendered Racism and the Production of Cultural Difference: Media representations and identity work among 'immigrant youth' in contemporary Sweden." *NORA - Nordic Journal of Feminist and Gender Research* 11(2): 78–88. https://doi.org/10.108%8038740310002932

Brugiavini, A., Cavapozzi, D., & Pan, Y. 2018. "Urban-rural differences in social policies: The case of the hukou system in China." In R. Taylor & J. Jaussaud (Eds.), *China's Global Political Economy: Managerial Perspectives* (pp. 55–66). Oxford: Routledge. https://books.google.com/books?hl=en&lr=&id=5mEPEAAAQBAJ&oi=fnd&pg=PA55&dq=hukou+system+and+secondary+education&ots=1NE-VCX5ZT&sig=_sZcBXNBXqeVo7Y4pU45d-Qr-NU#v=onepage&q=hukou%20system%20and%20secondary%20education&f=false

Brzozowski, J.-A., Taylor-Butts, A., & Johnson, S. 2006. "Victimization and offending among the Aboriginal population in Canada." *Juristat: Canadian Centre for Justice Statistics*, 26(3), 1–30. http://library.mtroyal.ca:2461/collections/Collection-R/Statcan/85-002-XIE/85-002-XIE2006003.pdf

Churchill, W. 2004. "Genocide by Any Other Name: American Indian Residential Schools in Context." *Genocide, War Crimes and the West: History and Complicity*, ed. Adam Jones, (London: Zed Books), 1–15.

Connell Szasz, M. 1980. " 'Poor Richard' Meets the Native American: Schooling for Young Indian Women in Eighteenth-Century Connecticut." In *Historical Review* 49(2).

Deaton, B. J., Scholz, A., & Lipka, B. 2020. "An Empirical Assessment of Food Security on First Nations in Canada." *Canadian Journal of Agricultural Economics* 68(1): 5–19. https://doi.org/10.1111/cjag.12208

Dewey, J. 2008. *The Early Works of John Dewey*, Volume 5, 1882 - 1898: Early Essays, 1895-1898, Ed. J. A. Boydston, (Carbondale: Southern Illinois University Press.)

Dunaway, W. A. 2016. "Challenging the Global Apartheid Model." *Journal of World Systems Research* 22(1): 16–22. https://doi.org/10.5195/JWSR.1

Foucault, M. 1976. L'Histoire de la Sexualité: la volonté de savoir. (Paris: Editions Gallimard).

Frost, K. (2019). "First Nations Sovereignty, Environmental Justice, and Degrowth in Northwest BC, Canada." *Ecological Economics* 162(April): 133–142. https://doi.org/10.1016/j.ecolecon.2019.04.017

Hamilton, A. 1788. No. 68: The Mode of Electing the President. *The Independent Journal*.

Hammond, C., Thomas, R., Gifford, W., Poudrier, J., Hamilton, R., Brooks, C., Morrison, T., Scott, T., & Warner, D. 2017. "Cycles of Silence: First Nations Women Overcoming Social and Historical Barriers in Supportive Cancer Care." *Psycho-Oncology* 26(2): 191–198. https://doi.org/10.1002/pon.4335

Hu, B., & West, A. (2014). "Exam-Oriented Education and Implementation of Education Policy for Migrant Children in Urban China." *Educational Psychology*, 41(3): 249-267. https://doi.org/10.108%3055698.2014.977780

Hu, M., Daley, A., & Warman, C. 2019. "Literacy, Numeracy, Technology Skill, and Labour Market Outcomes among Indigenous Peoples in Canada." *Canadian Public Policy* 45(1): 48–73. https://doi.org/10.3138/cpp.2017-068

Hung, J. 2022. "Hukou System Influencing the Structural, Institutional Inequalities in China: The Multifaceted Disadvantages Rural Hukou Holders Face." *Social Sciences* 2022 11(5):194. https://doi.org/10.3390/SOCSCI11050194

Ingram, L.-A. 2013. *Citizen-girls: Girls' perspectives on gender, citizenship and schooling*. University of Toronto.

Lewis, T. L. 2017. Spiritual Mothers Educating Republican Daughters: Elizabeth Bayley Seton, The Sisters of Charity, and St. Joseph's Academy for Girls. Houston State University.

Luo, J., & Wang, X. 2020. "Hukou Identity and Trust: Evidence from a framed field experiment in China." *China Economic Review* 59: 101383. https://doi.org/10.1016/J.CHIECO.2019.101383

MacDonald, D. B. 2015. "Canada's History Wars: Indigenous genocide and public memory in the United States, Australia and Canada." *Journal of Genocide Research* 17(4): 411–431. https://doi.org/10.1080/14623528.2015.1096583

Maina, N. K. 2006. "Interpretation of the Qur'an and its Influence on Girls' Access to Secondary School Education in Kenya." In *Contemporary Developmental Issues in Kenya*, eds. John Shiundu, Paul Achola, Helen Mondoh, Grephas Opata, Billy Ng'ong'ah (Egerton, Kenya: Organisation for Social Science Research in Eastern and Southern Africa) .

Meyer, Heinz-Dieter, & Benavot 2014. *PISA, Power, and Policy: the Emergence of Global Educational Governance*. (Oxford: Symposium Books).

Milloy, J. S., & McCallum, M. J. 2017. *A National Crime: The Canadian Government and the Residential School System* (Anniversary). University of Manitoba.

Mohanty, C. T., Russo, A., & Torres, L. 1991. *Third World Women and the Politics of Feminism*. (Bloomington: Indiana University Press).

Morris, E. W. 2007. "Ladies" or "Loudies"? *Youth & Society* 38(4): 490–515. https://doi.org/10.1177/0044118X06296778

Mujuni, J. B. 2015. *A Case Study of Cooperative Learning in Bushenyi District in Uganda: Educational leaders' and teachers' perceptions*. (Chicago: Loyola University).

Mullis, I., Martin, M., Foy, P., & Hooper, M. 2017. *TIMSS 2015: International Results in Mathematics*. http://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/full%20pdfs/T15-International-Results-in-Mathematics.pdf

Romanus Pontifex. 1455. https://www.papalencyclicals.net/nichol05/romanus-pontifex.htm?fbclid=IwAR1S2d6t4KTxVtWasYcNAGMMTOPDPcwUazdvG_S9chRIAxEHaDCIkdDtOL4

Palmater, P. 2022. Pope Francis, we don't accept your hollow apology. Here's why | The Star. *Toronto Star*. https://www.thestar.com/opinion/contributors/2022/07/26/pope-francis-we-dont-accept-your-hollow-apology-heres-why.html

Paradis, D. 2022. Reaction from survivors mixed as Pope Francis apologizes. *APTN*. https://www.aptnnews.ca/national-news/reaction-from-survivors-apology-pope-francis-maskwacis-alberta/

Paxton, K. A. 2006. "Learning Gender." In *Boarding School Blues: Revisiting American Indian Educational Experiences*. Lincoln: Bison Books (University of Nebraska Press).

Rostam-kolayi, J. 2008. "Origins of Iran's Moderns Girls' Schools: From Private/National to Public/State" *Journal of Middle East Women's Studies* 4(3), 58–88.

Sasaki, F. 2017. "Does Internet Use Provide a Deeper Sense of Political Empowerment to the Less Educated?" *Information Communication and Society* 20(10): 1445–1463. https://doi.org/10.1080/1369118X.2016.1229005

Sekamwa, J. 2000. *History and Development of Education in Uganda*. (Kampala, Uganda: Fountain Publishers).

Smylie, J., Fell, D., & Ohlsson, A. 2010. "A Review of Aboriginal Infant Mortality Rates in Canada: Striking and Persistent Aboriginal/Non-Aboriginal Inequities." *Canadian Journal of Public Health* 101(2): 143–148.

Spearman, J., & Watt, H. M. G. 2013. "Perception Shapes Experience: The Influence of Actual and Perceived Classroom Environment Dimensions on Girls' Motivations for Science." *Learning Environments Research* 16(2): 217–238. https://doi.org/10.1007/s10984-013-9129-7

Tan, Y., Kwan, M. P., & Chai, Y. 2022. "How Chinese Hukou System Shapes Ethnic Dissimilarity in Daily Activities: A Study of Xining, China." *Cities* 122: 103520. https://doi.org/10.1016/J.CITIES.2021.103520

Trennert, R. A. 1982. "Educating Indian Girls at Nonreservation Boarding Schools, 1878-1920." *Western Historical Quarterly* 13(3): 271–290.

Turyasiimwa, A. 2020. *Analyzing Policy and Practice in the Education of the Girl-Child in Uganda*. PhD diss., California Lutheran College.

Convention on the Prevention and Punishment of the Crime of Genocide, Pub. L. No. UNGA Resolution 260 A (III). 1948.

Watts, R. 2013. "Society, Education and the State: Gender perspectives on an old debate." *Paedagogica Historica* 49(1): 17–33. https://doi.org/10.108‰309230.2012.745886

Yuval-davis, N., Lee, L., 2009. "Interview with Professor Nira Yuval-Davis: After Gender and Nation." *Studies in Ethnicity and Nationalism* 9(1): 128–138. https://doi.org/10.1111/j.1754-9469.2009.01034.x

Zhang, D., Li, X., & Xue, J. 2015. "Education Inequality between Rural and Urban Areas of the People's Republic of China, Migrants' Children Education, and Some Implications." *Asian Development Review* 32(1): 196–224. https://doi.org/10.1162/ADEV_A_00042

Zhou, S., & Cheung, M. 2017. "Hukou System Effects on Migrant Children's Education in China: Learning from past disparities." *International Social Work* 60(6): 1327–1342. https://doi.org/10.1177/0020872817725134

ElizabethAnn S. Kelly, University of Albany