A leadership challenge paper in support of UNESCO Futures of Education 2050

Mobilizing Artificial Intelligence to Re-imagine Large-Scale Assessments

Pirasanya Ganasuntharam

Department of Leadership, Higher and Adult Education, Ontario Institute for Studies in Education

pirasanya.gnanasuntharam@mail.utoronto.ca

Technology has had a transformative impact on education since its introduction into teaching and learning spaces during the Cold War era (Christensen, 2019). Ed-tech companies and private sector funding have since shaped and re-defined how teaching, curriculum, pedagogy, and assessment are conceptualized to nurture human potential, growth, and progress. This mutually beneficial relationship was catapulted to unprecedented heights by the COVID-19 pandemic, when 1.6 billion students transitioning to online learning presented an ideal environment for further digitization of educational spaces. Various stakeholders, including ed-tech companies, were motivated to adapt and optimize digital spaces to maintain standards and enhance learning.

This paper examines one aspect of this relationship: large-scale assessments. The strategic foresight tool of Causal Layered Analysis is applied to examine the pervasiveness of standardized testing regimes by reimagining the role of technology as an assessment tool to support a more collective and unified purpose of education. Specifically, it seeks to imagine a possible future where artificial intelligence is equitably utilized to ensure large-scale standardized testing programs are relevant and responsive to the cultural identities of its global participants.

Positioning High-Stakes Standardized Tests As Part Of Our Educational Future.

Zhao *et al* frames education as a wicked problem, one that requires "deciphering what matters in education and then what should be measured and ultimately how to measure" (Zhao et al, 2019, p 274). This becomes a difficult feat when understood within a complex ecologies lens that situates education as an institution influenced and impacted by intersecting systems, including historical and contemporary forms of colonialism, patriarchy, neoliberalism, and capitalism. Neoliberalism's contribution to education is the encouragement of competition and comparison between nations in pursuit of technical, political, economic, and socio-cultural progress and power (Addey & Sam, 2019, p8). One way nations have become ensnared in this web is through participation in international large-scale assessments (ILSAs) and national large-scale assessments (NLSAs).

High stakes standardized testing is deeply solidified in the fabric of our local and global educational culture due to high financial, social, economic, political, and emotional investments and gains. Namely, PISA, the OECD's Programme for International Student Assessment, often regarded as the gold standard of international testing, had nearly 80 nations participate in the rankings in 2018.

While ILSAs like PISA have claimed not to measure school knowledge or competencies, results have been used to determine the quality of national school systems, with score-dependent implications for national and global aspirations. "Empirical evidence shows that ILSA participation is driven by multiple rationales, and many of these go beyond educational purposes to include political and economic purposes" (Addey and Sellar, 2018, p6). Politically, scores have been used to further international relations and generate evidence for policy mandates that favour certain agendas. For example, the United States' testing movement was initiated by the 1983 publication of A Nation at Risk, and continues to persist through politically-funded laws and policies like the Bush Administration's No Child Left Behind Act (2002) and the Obama Administration's Race to the Top initiative (2009). Countries have also used results to develop human capital for economic growth and further international aid-dependent agendas. *Edu-business* has also become a booming global market, often fueled by the results of the large-scale international studies (Sjoberg, 2019, p41). Pearson, the largest education corporation, boasted a test-making profit of $9 billion in 2017.

Ground-level inequities and research point to the dangers of over-reliance on ILSAs and NLSAs to shape educational policy and development at the cost of authentic learning and knowledge attainment. However, ILSAs and NLSAs continue to persist as an ever-expanding presence, that can be understood through the interactions of particular litanies, systems, and metaphors illustrated by Causal Layered Analysis (Inayatullah, 2017). The following illustrates how policymaking and policy actors have mobilized large-scale assessments within the context of particular narratives of what constitutes a desired future.

LITANY	● "when done right, standardized tests (STs) really do reflect what a student knows" ● "STs accurately assess students based on measured abilities" ● "STs are objective measures of aptitude and ability" ● "STs offer a fair and equal shot at educational, social, and economic achievement" ● "STs level the playing field through objective and competitive rankings"
SYSTEMIC CAUSES	● multi-national education corporations and the financial profitability of testing ● economic development (both national and international) ● economic expansion of top scoring countries into developing nations through aid. ● political gains through results-dependent policy-making agendas
WORLDVIEWS	● meritocracy (hard work=success) ● universalism (objective, empirical, evidence-based science) ● globalization ● neoliberalism (competition=progress and efficiency) ● individualism ● "survival of the fittest" (eliminating "inferior" schools through natural selection)
METAPHORS AND MYTHS	● "accountability through ST is key to teacher motivation" ● "STEM is the way of the future" ● "it is common sense that high scores in science and mathematics are good predictors of future economic prosperity" ● "education is an investment in the future of the country's economy" ● "human resources are the prime drivers in the modern economy" (human capital theory) ● "flexible and marketable workforce is the driving force of progress in the 21st century"

Problematizing Standardized National and International Assessments of Students

While large-scale assessments like PISA are widely believed to measure the preparedness level of young adults in meeting the challenges of the future, and if whether they have the capacity to become life-long learners, considerable limitations need to be addressed as efforts to renew and broaden their scope are pursued (Sjoberg, 2019, p17). PISA, for example, measures 15-year-olds' abilities to apply their reading, mathematics, and science knowledge and skills to meet real-life challenges. Its development is based on a bidding structure to select subject-matter experts and psychometricians who set standards based on politically aligned frames using English as their primary mode of communication. This problematically establishes and projects a particular Euro-centric worldview as a normative measurement tool to evaluate the quality of all participating countries' education systems. It produces a "winning set of outcomes [that] become codified as curriculum standards, accountability measures for schools and teachers, and bases for high-stakes decisions about the life of students " (Zhao et al, 2019, 268). In order to re-conceptualize a future where *what matters* include standardized tests that are equitably constructed, administered, measured, and applied, two critical perspectives need to be examined: Macro (claim of universality) and Micro (impact of ILSAs and NLSAs on student well-being and mental health).

On a macro-scale, universalism is a philosophical belief in a fundamental truth (Sjoberg, 2019, p21). The post-structural dilemma with universalism is that it does not exist. It is complicated by socio-cultural understandings of knowledge production and mobilization. Since birth, humans are heavily influenced by modes of socialization to observe, engage, make sense of, and navigate the world around us. This interaction constructs non-neutral forms of knowledge, informed by our cultural nuances, worldviews, biases, and mindsets, "where there is no one profile of qualities that is universally applicable to all tasks, jobs, and professions" (Zhao et al., 2019, p269). Set within this context, Sjoberg problematizes PISA's claim of its results examining the preparedness of young people for adult life by asking whether "it is possible to measure the quality of a country's education by indicators that are common (universal), independent of school systems, social structure, traditions,

culture, natural conditions, ways of living, modes of production, etc" (Sjoberg, 2019, p21). He complicates Enlightenment era principles of rationality, objectivity, and the Scientific Method, set within a climate of imperialism that positioned Eurocentric values as neutral while otherizing other ways of knowing as inferior and backwards. Thus, a claim that presumes to measure preparedness through common indicators becomes an imposition of a Eurocentric and monolithic view of societies on the rest of the world. Zhao *et al.* refers to this constrictive approach as a wicked problem, where "measurement is often treated as a tame technical problem that can be solved following traditional linear, analytic approaches" (Zhao et al, 2019, 264). This linear approach promotes a sense of "sameness" in how assessments are developed, implemented, and analyzed that polarizes the collective and the individual, without acknowledging that the collective identity encompasses unique, individual differences - one that embraces different skills and competencies based on the customs, values, and societal needs of each individual. It is imperative to centre student individuality and harness this power through the collective in order to authentically measure *what matters.*
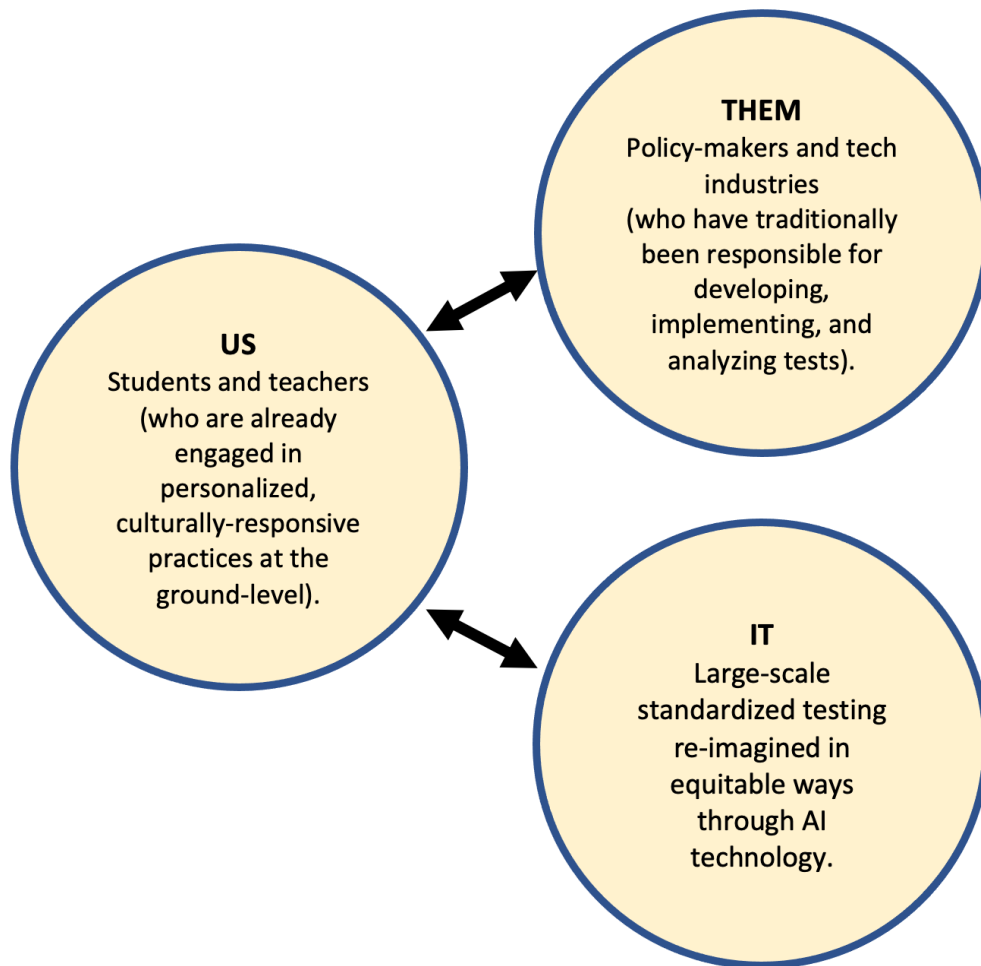
On a micro-scale, the ground-level impacts of ILSAs and NLSAs on student well-being and mental health are troubling. Testing marginalizes those students who learn differently, with learning preferences not tested through "pencil and paper" style formats. Skerrett and Hargreaves have noted that "standardization and high-stakes testing have contributed to the institutionalization of inequities in both Canadian and American education systems, resulting in the overrepresentation of minority/racialized students in lower academic tracks" (Rezai-Rashti and Bob Lingard, 2020, p4). It also targets students who have special needs, are English Language Learners, recent immigrants, low-income, and Indigenous learners who draw upon non-Eurocentric funds of knowledge and capital, and may not have the required skills and competencies to write a time-sensitive and space-restricted exam (Eizadirad, 2019, p24). This generates testing-induced stress and anxiety and an internalized feeling of shame and deficit-thinking linked to failure. Kearns research on youth voice related to failure on the Ontario Secondary School Literacy Test found that "some youth not only expressed "shock" and lack of understanding at the test results, but felt "shame," "degraded," "humiliated," "stressed," "a little less smart," "like a loser," and

expressed "fear," upon learning that they had failed" (Kearns, 2011, p119). These sentiments run contrary to EQAO's policy promise of using assessment results to pursue excellence and equity for every student in the Ontario school system.

Actualizing UNESCO's 2050 Collective Vision For An Inclusive Future

Acknowledging students' individual subjectivities and diverse ways of interacting with the world based on their cultural worldviews is critical to ensuring an equitable future. According to UNESCO's 2050 *Learning to Become* initiative, actualizing this vision is a collective endeavour towards building a future where every student has equitable access to opportunities and outcomes that enable them to thrive and be contributing democratic citizens to their local and global communities.

This realization problematizes current large-scale standardized testing frameworks in the name of accountability. What do we need to be accountable for to realize that "knowledge and learning are humanity's greatest renewable resource for responding to challenges and inventing alternatives" (UNESCO, 2019)? Certainly, this form of accountability cannot be actualized through ILSAs or NLSAs that promote mechanistic models of testing knowledge in primarily three domains: science, mathematics and reading, using rigid 'pencil and paper' formats. If the purpose of schooling, as a key socializing agent, includes the development and enhancement of non-cognitive values of well-being, citizenship, solidarity, compassion, empathy, curiosity, critical thinking, and mental health, then *learning to become* needs to extend beyond preparing efficient citizens to include their personal, human, and social development as responsible citizens. This requires the application of artificial intelligence driven by a commitment to the well-being of all students. The following illustrates the promise of applying AI in a way that avoids the marginalization and 'othering' of students adapting the construct from Godet (1987) that represents this commitment to inclusivity.

THEM
Policy-makers and tech industries (who have traditionally been responsible for developing, implementing, and analyzing tests).

US
Students and teachers (who are already engaged in personalized, culturally-responsive practices at the ground-level).

IT
Large-scale standardized testing re-imagined in equitable ways through AI technology.

Reimagining A Radical Future Where High-stakes Assessments Advance Equity in Education

Mobilizing UNESCO'S 2050 *Learning to Become* initiative to disrupt an assessment culture that presents education as a set of skills that can be measured, compared, and ranked requires educational stakeholders to re-vision and transform learning spaces and knowledge acquisition by "leveraging humanity's collective intelligence" (UNESCO, 2019).

Traditionally, the development of ILSAs and NLSAs have primarily involved consultants, research think-tanks, test providers, and policy-makers. *What* to measure has involved top-down authoritative decision-making by policy and academia experts, while *how* to measure the *what* has imbued a competitive environment, with ed-tech companies like Pearson vying for control and dominance (Zhao et al, 2019, p267). Transformational change,

however, necessitates re-imagining who is at the table? Who is missing from the table? And whose critical voices are silenced or absent? This opens possibilities for having youth, teachers, and educational leaders join the conversation and develop testing materials that incorporate diverse ways of knowing and interacting with the world. Zhao et al frames this as a collaborative approach that must involve "students, parents, and teachers - the three groups of stakeholders with the most at stake in the solution - [but] have traditionally played only marginal roles" (Zhao et al, 2019, p268). In particular, teacher autonomy in decision-making must be valued through this collaboration to avoid powerful technological take-overs that have the potential to make the teaching profession obsolete through cost-effective innovations and mechanisms (Sellar, 2020).

Applying Michel Godet's 'Us-Them-It' model, human agency is elevated when the 'US', primarily teachers and students, work collectively with the 'THEM', policy-makers and tech industries, to develop, implement, and analyze high stakes tests, in pursuit of the wandering 'IT', standardized testing re-imagined in equitable ways through artificial intelligence technology. This is a critical step, especially when acknowledging that there "is virtually no literature which engages students' perspectives" because it is assumed that "assessments have minimal impact on [youth] subjectivities or that youth concerns…are merely a backdrop to the assessment process" (Wiliam, 1999, p344–345). Then the question becomes: in the age of big data and technological advancements, how might one go about attaining this future using artificial intelligence?

The OECD's launch of Education 2030 at the 2019 Vancouver ed-policy conference that referenced the promise of artificial intelligence replacing teachers and schools as subject-based transmitters of knowledge, with teachers becoming exclusively responsible for socio-emotional aspects of development, is one way of conceptualizing AI's influence on education. China's AI-directed tutoring service, Squirrel, also promotes adaptive learning by aiming to employ sophisticated algorithms to pinpoint exactly what students know and don't know through standardized, skills-based knowledge acquisition programs. However, Harvard professor Chris Dede positions adaptive learning as distinct from personalized learning, which "pays attention to what [students] want to know and how they learn best"

(Hao, 2019). Using this personalized application of AI technology, this paper positions AI *before* classroom-based subject learning even begins.

Sophisticated AI technology will track student activities and behaviours to compute data about individual learning preferences, multiple intelligences, and cultural knowledges. In this regard, the application of analytics to the field of education is for measuring, collecting, analyzing, and reporting data *about* learners and their contexts, for purposes of *understanding and optimizing* learning and the environments in which it occurs (Ferguson, 2012, p305). Adapting OCAD Professor Treviranus's personalized learning levels of path and destination to AI-infused ILSAs and NLSAs, their applicability to student success can be seen in the following way:

Path = students given differentiated questions based on AI tracking of preferences, intelligences and cultural nuances to arrive at the same objectives being tested

Destination: big data (gathered through differentiated ILSAs and NLSAs) analyzed in highly personalized ways by teachers and school districts to provide students with diverse learning opportunities to pursue goals because it "takes their interests and needs into account to orchestrate the motivation and time for each student so they are able to make progress" (Hao, 2019).

When this process is done ethically, with sensitivities paid towards privacy of student data, it would potentially be revolutionary in using student voice as a data collection tool to inform test content creation. This big data will then be used in conjunction with other forms of data and expertise on regional differences of countries and cultures to support educators and ed-companies develop materials that test the same content knowledge, but through culturally relevant and differentiated scenarios and applications. Following is an example of a standardized test question that has been contextualized to fit each student profile:

STUDENT A	STUDENT B
AI-generated data: ● Identity: Indigenous (\North West Territories, Canada) ● Multiple Intelligences (based on AI tracking): Spatial, Bodily-Kinesthetic, Naturalist ● Highly values land-based, sustainability practices	**AI-generated data:** ● Identity: Japanese ● Multiple Intelligences (based on AI tracking): Logical/Mathematical, Musical, Intrapersonal ● Highly values human connections, technological advancements, and spiritual traditions
Content knowledge tested: proportional reasoning **Assessment question (computer, pencil/paper)**: Visually build a fish rack using dry salmon to demonstrate understanding of proportional reasoning based on relative units of measure. Idea is for the student to construct a fish rack proportional to the family member who will use it, and demonstrate their understanding by applying this context to Western standard units of measurement. *Fishing helps connect Indigenous peoples to their environment, with important cultural, social, and economical implications and values. This scenario also decolonizes approaches to assessment.	**Content knowledge tested**: proportional reasoning **Assessment question (computer, pencil/paper)**: Akira and Yuto ride the Shinkansen (Japanese Bullet Train) to school everyday. If it takes Akira 30 minutes to drive 50 kilometres, calculate how long it will take Yuto to ride 100 kilometres to demonstrate understanding of proportional reasoning based on relative units of measure.
Questions demonstrate knowledge and competencies in critical thinking skills, ownership of learning, and sense-making of mathematical concepts through differentiated questions contextualized within cultural ways of being and knowing.	

Successful implementation and application of results require a high level of trust. One solution is to re-think how we incorporate educator voice in the dissemination and utilization of results for best practices. Typically, results are shared with classroom teachers, with either no commitment for change or punitive repercussions for low results, based on the country and/or district's policy on uses of data results for system improvement. This data seeks instead to increase teacher motivation through incentivized

commitments (e.g. school-based and regional professional development workshops, departmental collaborations) to analyze student results and determine assets and gaps in performance due to inequities within each school and school district. This will not only re-focus efforts on those students not mastering foundational skills within a culturally responsive data environment, but also utilize results to identify instructional practices to target students requiring additional supports in closing the gap between those who are meeting curriculum expectations and those who are not. It is important to note that teachers will already have a socio-cultural understanding of their students based on strong relationships and responsive cultural practices. The results would serve to strengthen their instructional and pedagogical practices by providing additional dimensions for consideration.

Acknowledging the Risks of Artificial Intelligence

While AI mediated large-scale assessments are promising as a possible future, one would be remiss not to surface the risks involved in introducing such a transformative entity into the world of children and data collection. In fact, accounting for risks will allow for educational stakeholders to pre-emptively troubleshoot and mitigate potential shortcomings. One solution to preventing computational bias is through robust collaboration with teachers to ensure algorithms accounting for cultural nuances and sensitivities represent a diverse array of lived identities. It is also of critical importance to develop AI technology that accounts for student privacy concerns. Confidentiality of student information when employing any data-gathering technology is important, especially those related to culture and identity. This sophistication is necessary to avoid data collection and analysis informed by cultural biases and generalizations.

Conclusion

Artificial Intelligence technology provides an opportunity to explore possible futures for ILSAs and NLSAs, which have long been criticized by preeminent scholars and researchers as unethical entities forsaking values of human connectivity and socio-emotional well-being for neoliberal values of competition, progress, and efficiency. Utilizing artificial intelligence in culturally relevant and responsive ways calls for radical change to existing testing mechanisms that deliver the same content to all students, irrespective of regional

and cultural differences. It highlights both the existential and practical preponderance of the collective versus the individual. The assessment of individual students is...one of the key ways that we understand, measure and compare - through accountability systems – the progress of schools...and the health of our whole education system. (Baker, Smith & Anissa, 2019, p41). If AI can be used to shift ILSAs and NLSAs from advancing standardized adaptive technology to culturally differentiated testing in personalized ways, the future holds great potential for how AI-informed big data can be used to meaningfully operationalize an educational system where individual differences are valued through the collective, such that learning and knowledge become the collective responsibility and right of all citizens.

References

Addey, C., & Sellar, S. (2019). Is it worth it? Rationales for (non) participation in international large-scale learning assessments. *Education Research and Foresight Working Papers.* UNESCO, Paris.

Agasisti, T., & Bowers, A. J. (2017). Data analytics and decision making in education: towards the educational data scientist as a key actor in schools and higher education institutions. *In Handbook of contemporary education economics.* Edward Elgar Publishing.

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges.

Christensen, D. (2019, August 13). The history of the emergence of technology in education. Retrieved November 10, 2020, from https://www.classcraft.com/blog/features/the-history-of-the-emergence-of-technology-in-education/

Godet. M. 1987. *Scenarios and Strategic Management*. Butterworth Scientific, London.

Goli M. Rezai-Rashti & Bob Lingard (2020). Test-based accountability, standardized testing and minority/racialized students' perspectives in urban schools in Canada and Australia. *Discourse: Studies in the Cultural Politics of Education.*

Hao, K. (2019). China has started a grand experiment in AI education. It could reshape how the world learns. *MIT Technology review.* Luettu, 1, 2019.

Inayatullah, S. (2017). Causal layered analysis. *Prospective and Strategic Foresight Toolbox*, 1-21.

Kearns, L. L. (2011). High-Stakes Standardized Testing & Marginalized Youth: An Examination of the Impact on Those Who Fail. *Canadian Journal of Education*, 34(2), 112-130.

Kertes, T. (n.d.). Why the OECD's education agenda matters to students, communities, and teachers in British Columbia, Tom Kertes, Teacher in Prince Rupert, BC. Retrieved July 21, 2020, from *https://bctf.ca/publications/ResearchReports.aspx?id=55284*

Mathison, S., & Ross, E. W. (Eds.). (2004). The nature and limits of standards-based reform and assessment: Defending public schools. *Teachers College Press.*

Sellar, S. (2020, October 20). "Reimagining the futures of assessment with teacher professional autonomy". Retrieved November 09, 2020, from *https://www.worldsofeducation.org/en/woe_homepage/woe_detail/17011/%E2%80 %9Creiagining-the-futures-of-assessment-with-teacher-professional- autonomy%E2%80%9D-by-sam sellar.*

Sjøberg, S. (2019). The PISA-syndrome–How the OECD has hijacked the way we perceive pupils, schools and education. *Confero: Essays on Education, Philosophy and Politics*, 7(1), 12-65.

UNESCO. (2020). UNESCO's Futures of Education. Retrieved July 20, 2020 from *https://en.unesco.org/futuresofeducation/initiative.*

Zhao, Y., Wehmeyer, M., Basham, J., & Hansen, D. (2019). Tackling the wicked problem of measuring what matters: Framing the questions. *ECNU Review of Education*, 2(3), 262-278.