International Journal of Educational Development 73 (2020) 102149

Contents lists available at ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev

Human Capital Index and the hidden penalty for non-participation in ILSAs

Ji Liu[a,*], Gita Steiner-Khamsi[b]

[a] School of Education, Shaanxi Normal University, PR China
[b] Teachers College, Columbia University, USA

ARTICLE INFO

Keywords:
Human Capital Index
International Large-Scale Assessments
Education policy
Development
Comparative education
Testing and accountability

ABSTRACT

The World Bank's Human Capital Index (HCI) aims to provide new information regarding future productivity of each country's workforce, by synchronizing available International Large-Scale Assessment (ILSA) and regional test program results. Linking the literature on ILSA participation, this study questions the problematic nature of this approach and revisits the comparability issue of ILSA results. We find that education systems are imposed upon a score penalty depending on which ILSA or regional test program they choose to partake in. In particular, our results show that (i) test-overlap systems used in the score synchronization procedure are systematically different from systems that only choose to participate in one ILSA or exclusively in regional tests, (ii) inter-test score exchange rate is volatile due to sampling design and cohort effects, (iii) test participation type alone accounts for about 57.8 percent of the variation in synchronized scores, and the score penalty is especially salient for systems that exclusively participate in regional test programs; the majority of which are low-income and lower-middle income countries. Findings in this study show how various intra- and extrapolations to compensate for missing data in effect introduce large score penalties for systems that either did not participate or only partially participated in ILSAs. Finally, this study contributes to research on reasons for participation in ILSAs and the global rise of test-based accountability reform, under which the World Bank's new HCI may be seen as a tool to incentivize participation in ILSAs by penalizing those governments that have chosen alternative, non-standardized paths for measuring learning outcomes of students.

1. Introduction

Economists have long been interested in studying human capital, and this enthusiasm was propelled forward in part by the theoretical foundations laid out by Schultz (1961) and Becker (1962), and coincided with the expansion of the global development project since World War II and more recent donor interests in education interventions (Chabbott, 2007; Heyneman, 2003; Shields and Menashy, 2019; Steiner-Khamsi, 2006). To date, there have been various attempts to conceptualize and quantify human capital, for instance, United Nations Development Programme (UNDP) has been routinely measuring and releasing the Human Development Index since early 1990s, whereas Barro and Lee (1993) are among the first academics to construct a cross-nationally comparable educational attainment dataset using census and household survey. Until only recently, the focus on human capital measurement has been primarily confined to education quantity measurements that rely on enrollment or attainment data. Notwithstanding, the global growth of International Large-Scale Assessments (ILSAs) and regional student testing programs has prompted recent attempts to examine student learning outcomes more closely (see Barro and Lee, 2001; Cohen and Soto, 2007; Hanushek and Kimko, 2000; Hanushek and Woessmann, 2012; Heyneman and Lee, 2012).

As part of its global call-to-action to invest more in people, the World Bank released its first Human Capital Index (HCI) results in October 2018, ambitiously covering 157 economies and close to 98 percent of the world's population (World Bank, 2018a). While the index captures public health information such as infant and adult mortality, its main novelty according to the World Bank (2018b, p. 4), lies in its ability to provide "direct measure of school quality and human capital." To this end, the World Bank (2018b) envisions the index's education component as a new macro indicator to help countries keep track of progress in education quality or the lack thereof. As indicated in several key publications, the World Bank (2018a, 2018b, 2018c, 2018d, 2019a) plans for the index to have far-reaching global impacts, and indicated its future priority in setting HCI as a central metric in education planning, public financing, and government accountability. More specifically, the Bank plans on utilizing HCI to strategically promote new measurement initiatives, inform lending programs, and address shortcomings in political incentives (World Bank, 2018a, p.10).

While the HCI holds a grand promise, there are many intricacies

* Corresponding author at: Shaanxi Normal University, 199 Chang'An South Ave., Tin Ka Ping Education Building, Xian City, Shaanxi Province, 710062, PR China.
E-mail addresses: jiliu@snnu.edu.cn (J. Liu), gs174@tc.columbia.edu (G. Steiner-Khamsi).

https://doi.org/10.1016/j.ijedudev.2019.102149
Received 21 July 2019; Received in revised form 28 October 2019; Accepted 16 December 2019

worth examining in detail. Importantly, the key education component in the HCI – Harmonized Learning Outcomes (HLO) – is extrapolated from various ILSAs and regional test programs by relying on an inter-test exchange rate that is calculated using test-overlap systems for given pairs of test programs. However, differences in test design, timing, sampling, non-school factors can substantially limit the usefulness of such an exercise. To illustrate, the World Bank (2018b) imputes HLO scores for Lesotho, which did not participate in any ILSA programs, using an inter-test exchange rate that is calculated with Botswana's scores. However, Lesotho and Botswana are far from similar; for instance, the per capita Gross Domestic Product (GDP, 2017 PPP) amount in Botswana is $17,024, or 5.5 times more than that of Lesotho ($2932) (World Bank, 2019b). The obvious leap of faith is that this inter-test exchange rate is country-invariant and transcends borders and contextual factors such as demographics, income, culture, etc. As such, there are several important implications that arise with centering ILSA score harmonization at the core of the new Human Capital Index. Firstly, the reliance on synchronizing ILSA results to produce an HLO score implies that countries who choose not to participate in any ILSA or regional test programs will not receive a score. It is conceivable that a score penalty, which can substantively limit future policy, financing, and programming prospects, will act as policy levers in pushing countries towards reconsidering ILSA non-participation decisions. Secondly, because the core component of HCI involves test result conversion based on test-overlap systems, therefore score conversion rates could be biased upwards, neutral, or downwards depending on which ILSAs or regional test program and which year a system chooses to participate in. As the expansive literature on ILSA participation and comparability has shown (see Addey et al., 2017; Addey and Sellar, 2019), the political, economic, and social costs often can both inhibit or incentivize participation depending on country context, which further amplifies the comparability issue inherent to ILSAs through inter-test score synchronization exercises.

More specifically in this study, we examine the 'penalty' complex in further detail, and address the following research questions: (i) How reliable is the intra-test exchange rate, which depends on the comparability of test-overlap systems? (ii) Does participation in different test programs matter for Harmonized Learning Outcome (HLO) results, and how much is the associated penalty? In this investigation, we argue that the introduction of HCI as a new education quality tracking device, centering on synchronizing ILSA scores, should be critically examined. In particular, the algorithmic procedures devised to impute the index adds a new dimension of complexity – penalties – to the widely documented 'reasons for ILSA engagement' literature, in that systems are penalized for both non-participation and partial-participation in ILSA benchmarking. While existing studies have explored in depth the demand-side reasons why countries choose to participate in ILSAs, our current study contributes to understanding new supply-side intricacies that are introduced through HCI calculations. By examining the non-participation and partial-participation effect, we show how extrapolations to compensate for missing data in effect introduces considerable score penalties for countries that either did not participate or only partially participated in ILSAs.

2. Coercive rationales for ILSA engagement: a focus on the supply-side

The spectacular growth of education systems participating in ILSAs is noticeable and deserves theorizing. In this list, some of the most influential ILSAs include *Programme for International Student Assessment* (PISA), *International Reading Literacy Study* (PIRLS), *Trends in International Mathematics and Science Study* (TIMSS), and regional assessments such as *Latin American Laboratory for Assessment of Quality in Education* (LLECE), *Southern and Eastern Africa Consortium for Monitoring Educational Quality* (SAQMEC), *Early Grade Reading Assessment* (EGRA), and so on (Fischman et al., 2019). In terms of PISA alone, 43 countries participated in 2000, 72 economies in 2015, and 80 economies in 2018. In analyzing the explosive growth of PISA and other ILSAs, researchers have proposed several explanations, ranging from broad ones, such as globalization and political pressure to be part of a larger international educational space, to very concrete ones, such as an ever-increasing number of evidence-driven policy actors who rely on international comparison for measuring the quality of their educational system (see Addey et al., 2017; Engel, 2015; Liu, 2019; Verger et al., 2019). It is noticeable that this body of research has focused on national governments and explored why they buy into ILSAs, that is, explain the rationales for their participation.

The attractiveness of ILSAs is in part related to the quantification of learning outcomes. Wendy Espeland (2015) and Radhika Gorur (2016) masterfully observe the advantages of numbers over complex narratives because one may attach one's own narratives to numbers. What is especially appealing to policy actors are *Organization for Economic Co-operation and Development* (OECD) and *International Association for the Evaluation of Educational Achievement* (IEA) studies, is the capacity to generate statistics, scores, ranking, and benchmarks that are based on international comparison or on comparison over time. Espeland (2015, p. 56) explains the dual process of *simplification* and *elaboration* involved in the usage of numbers. In a first step, numbers "erase narratives" by systematically removing the persons, institutions, or systems being evaluated by the indicator and the researcher doing the evaluation. This technology of simplification stimulates narratives, or as Espeland astutely observes:

> If the main job of indicators is to classify, reduce, simplify, to make visible certain kinds of knowledge, indicators are also generative in ways we sometimes ignore: the evoke narratives, stories about what the indicators mean, what are their virtues or limitations, who should use them to what effect, their promises and their failings. (Espeland, 2015, p. 65)

A few scholars have focused on the "narrative evoking" phase (Espeland, 2015, p. 65) of such studies and dissected what national governments interpret or project onto OECD reports or other international comparative studies based on their own policy context and agenda (Waldow and Steiner-Khamsi, 2019). Studies on the rationales for participation in ILSAs have typically attempted to address the following contradiction: a great number of governments chooses to continuously participate in ILSAs even though they have already learnt after the first round of participation about the strengths and weaknesses of their systems. Arguably, the focus on rationales for participation has generated two blind spots: (i) rationales for non-participation and (ii) strategies to ensure continuous participation. The first research area focuses on governments (i.e. Botswana, South Africa, Kyrgyzstan, Switzerland) that discontinued, or considered a discontinuation (i.e. China, Mexico, Vietnam), respectively, the participation in an ILSA and the second research area focuses on international organizations that develop, administer, and analyze ILSAs. For the second under-explored research area, OECD's strategies to ensure continued interest is worth mentioning briefly.

In regards to the first area mentioned above, more research is needed on *rationales for non-participation or discontinuation*, respectively. For instance, Kyrgyzstan discontinued its participation for the PISA 2012 round after scoring at the very bottom in PISA 2006 (out of 47 countries and economies) and PISA 2009 (out of 65 countries and economies). In disbelief over the poor results, the Ministry of Education and Science established its own national assessment center. Notwithstanding, not all "ILSA drop-outs" are education systems that score at the bottom of a league table (Addey and Sellar, 2019; Lockheed and Wagemaker, 2013). For instance, South Africa discontinued TIMSS for resource reasons (Wiseman, 2013) and Mongolia dropped out of TIMSS due to technical capacity constraints (Addey, 2015). League leaders also have good reasons to discontinue their participation in ILSAs. For example, Switzerland only participated in two rounds of