A Report on “Measuring Human Capital Using Global Learning Data” by Angrist et al. (2021)

Reviewer 2

February 13, 2026

v1

isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.

Plato, *The Apology of Socrates*, 21d

To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.

Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring Human Capital Using Global Learning Data. *Nature*, Vol. 592, No. 7854, pp. 403–408.

URL: https://doi.org/10.1038/s41586-021-03323-7

Abstract Summary: This article constructs a globally comparable database of learning outcomes for 164 countries from 2000 to 2017 to bridge the gap in comparable learning metrics, particularly in developing countries, and uses this data to estimate the role of human capital in explaining cross-country income differences. The database, referred to as the Harmonized Learning Outcomes (HLO) database, provides a measure of human capital that is more closely associated with economic growth than current measures.

Key Methodology: Construction of the Harmonized Learning Outcomes (HLO) database by linking international and regional student achievement tests using regression and linear linking methods, followed by development accounting and growth regressions.

Research Question: How can a globally comparable database of learning outcomes be constructed, and what is the role of human capital, measured by this new data, in explaining cross-country income differences and economic growth?

Summary

Is It Credible?

This article presents a significant effort to construct a "globally comparable database" of learning outcomes (HLO) for 164 countries, aiming to replace simple measures of schooling quantity with measures of quality. The authors claim this new metric reveals that human capital accounts for "a fifth to around half" of cross-country income differences and that learning is "more strongly associated with economic growth" than traditional indicators (p. 403). While the construction of the HLO database is a substantial logistical achievement, the credibility of the resulting "global" comparability and the magnitude of the economic claims are constrained by methodological limitations and strong assumptions.

The core of the article's contribution—the harmonization of regional and international test scores—relies on linking functions derived from a subset of countries. The validity of these links is particularly fragile for the developing regions that are central to the article's novelty. Sensitivity analyses in the supplementary materials reveal that for African regional assessments (SACMEQ and PASEC), using random subsets of countries to generate the link can alter scores by 19 to 25 points on a scale with a standard deviation of 100 (p. 9 of Supplementary Information). The authors acknowledge that this variation is consistent with low country overlap and advise that scores should be interpreted with caution, focusing on "relative ranks" rather than precise values (p. 13 of Supplementary Information). Furthermore, the linking functions assume that the relationship between regional and international tests remained constant over nearly two decades (p. 407). While the authors defend this as reasonable due to standardized testing practices since the mid-1990s, if the relative difficulty of these assessments shifted due to curriculum changes or psychometric updates, the reported trends—specifically the "flat" learning profiles despite rising enrollment—could be artifacts of the harmonization process rather than genuine educational stagnation.

The economic analysis utilizing this data also faces scrutiny regarding the magnitude of its claims. The assertion that a “1% in learning is associated with a change of 7.2% in annual growth” relies on a specific interpretation of the regression coefficient (0.072) in a log-level specification (p. 405). Depending on how the dependent variable (annual growth rate) is scaled, this could represent a small absolute percentage point increase rather than a large relative change in the growth rate itself. Additionally, while the growth regressions exclude countries with civil war or high resource rents to address some confounders, they omit other standard controls such as institutional quality or trade openness. It remains possible that underlying state capacity drives both higher test scores and economic growth. Similarly, the development accounting exercise, which attributes up to half of income differences to human capital, depends heavily on assumed parameters for returns to schooling and learning ($w = 0.20$) rather than parameters estimated from the data itself (p. 409). Consequently, these results should be viewed as illustrative simulations based on specific assumptions rather than empirical proofs of human capital’s contribution.

The Bottom Line

The HLO database represents a valuable step toward quantifying educational quality globally, but the claim of “global comparability” is weakened by instability in the linking methodology for developing nations. While the “learning crisis” narrative is likely directionally correct, the precise trends and the magnitude of the link between learning and economic growth are uncertain due to potential measurement artifacts and coefficient interpretation ambiguities. The data is best used for broad regional comparisons rather than precise country-level rankings or definitive causal economic modeling.

Potential Issues

Ambiguity in economic growth regression interpretation: The article's central claim about the economic significance of its Harmonized Learning Outcomes (HLO) measure relies on a specific interpretation of the regression model in Table 2. The text states, "We observe that a change of 1% in learning is associated with a change of 7.2% in annual growth" (p. 405). This is derived from a coefficient of 0.072 in a log-level model, where the independent variable (HLO) is log-transformed. If the dependent variable is the raw growth rate (e.g., 2.0 for 2%), a coefficient of 0.072 implies a negligible absolute change. The authors' phrasing suggests a relative change *in* the growth rate (e.g., moving from 2.0% to 2.14%), which assumes a specific scaling of the dependent variable not explicitly defined in the table notes. This ambiguity makes it difficult to assess whether the estimated economic impact is as substantial as claimed.

Omitted variable bias in growth regressions: The conclusion that the HLO measure has a "stronger association with growth than alternative human capital measures" is based on regression models that may suffer from omitted variable bias (p. 405). While the authors address some confounders by excluding countries with civil war or high resource rents, the cross-sectional models in Table 2 do not control for other standard determinants of growth, such as institutional quality, trade openness, or investment rates (p. 406). These factors are likely correlated with both a country's educational quality and its economic growth. By omitting them, the HLO coefficient may be capturing the effects of these unmeasured factors, making it difficult to attribute the observed association solely to learning.

Reliance on fixed and extrapolated parameters in development accounting: The development accounting exercise, which estimates that human capital accounts for "between a fifth to around half" of income differences, relies on a model with key parameters that are assumed rather than estimated from the data (p. 405). The return to learning (w) is based on a value of 0.20 derived from "US data," an assumption the authors acknowledge in the supplement may not be appropriate for the developing countries that make up two-thirds of the sample (p. 409). While the article includes a sensitivity analysis for this parameter, the return to schooling (r) is fixed at 0.10 across all specifications (p. 409). These strong assumptions about fixed, universal returns to schooling and learning directly determine the final estimates, making the exercise more of an illustrative simulation than an empirical finding.

Instability of the harmonization procedure for developing countries: The article's claim to create a "globally comparable database" is challenged by evidence in its own supplementary materials regarding the regional assessments (PASEC and SACMEQ) that provide data for many African nations. A sensitivity test reveals that using random subsets of countries to create the linking function can lead to average score differences of "19 points for SACMEQ and 25 points for PASEC," a substantial variation on a scale with a standard deviation of 100 (p. 9 of Supplementary Information). The authors attribute this to low country overlap and advise focusing on "relative ranks" (p. 13 of Supplementary Information). While the authors are transparent about this limitation in the supplement, the magnitude of this instability raises questions about the precision of the data for the very countries whose inclusion is a key part of the article's contribution.

Assumption of stable test relationships over time: The time-series analysis of learning trends rests on the assumption that the relationship between different standardized tests remains constant over the entire 2000–2017 period. The authors state, "In fixing the linking function, we assume that the relationship between tests stays constant across rounds," justifying this as reasonable due to the standardization of testing approaches since the mid-1990s (p. 407). However, curricula, testing frameworks, and psychometric properties can evolve over nearly two decades. If the relative difficulty between a regional and an international test has changed, this method would incorrectly attribute that change to student learning rather than to a measurement artifact.

Generalizability of linking functions: The functions used to convert regional test scores to an international scale are derived from the subset of "linking countries" that participated in both types of assessments. For example, the link for the Latin American test (LLECE) is derived from countries like Chile and Colombia (p. 21 of Supplementary Information). While the authors use all available overlapping countries to maximize the sample size, this approach assumes that the relationship between the tests observed in these countries is generalizable to their neighbors (p. 407). The article relies on this assumption of external validity without direct evidence that the test relationship holds for non-linking countries.

Potential for unresolved selection bias to distort trends: The article concludes that rising school enrollment has not translated into learning, but this finding may be affected by selection bias. As enrollment expands, more marginalized students typically enter the school system, which can mechanically depress average test scores. The authors acknowledge this "selection effect" and argue their estimates are a "conservative upper bound" because primary enrollment is already high (p. 14 of Supplementary Information). However, this does not fully resolve the issue of *differential* selection, where the rate of change in the test-taking population's composition differs across countries and over time, potentially creating spurious trends.

Methodological and reporting issues: Several methodological choices and reporting practices raise concerns. First, a key regression result in Table 2 (column 8) reports a p-value of 0.914 for the "Human capital (HDI)" variable, which is inconsistent with its coefficient of -0.028 and standard error of 0.022; the p-value should be approximately 0.21 unless specific clustering or robust error calculations not detailed in the table are driving the result (p. 406). Second, the exclusion of all mathematics scores from the PASEC assessment is justified only briefly in the supplement as being the "least reliable linking function," a decision that affects a key data source for Francophone Africa (p. 20 of Supplementary Information). Third, the methodology for a bespoke adjustment that substantially lowers China's learning score is not described in the article but is outsourced to a separate working paper (p. 20 of Supplementary Information). Finally, the analysis of trends in enrollment and learning from 2000 to 2015 relies on an assumption for a third of the period; as enrollment data was only available to 2010, the authors "make a conservative assumption that enrolment rates persist through 2015" (p. 408).

Confounding of student development and test difficulty: The methodology for linking some tests introduces a potential confounder by comparing student cohorts at different grade levels. The supplement notes this possibility: "...linking PIRLS 2001 grade 4 with SACMEQ 2000 grade 6 might capture a grade difference... in addition to difficulty" (p. 19 of Supplementary Information). A two-year gap in schooling represents a significant difference in cognitive development. The authors attempt to address this by citing a sensitivity analysis for EGRA that showed "small differences" (p. 19 of Supplementary Information). However, it is not demonstrated that this finding generalizes to other, more consequential cross-grade links, leaving open the possibility that the resulting linking function is a composite of both test difficulty and the developmental gap between student populations.

Exclusion of outlier countries: In the analysis comparing trends in enrollment and learning, the authors state they "omitted four countries (Mozambique, Niger, Cameroon and Benin) that are outliers... which can bias average cross-country trends" (p. 408). All four are in sub-Saharan Africa. While the authors exclude these countries precisely because they would distort the average trend, it is important to note that these "outliers" may represent genuine cases where rapid enrollment expansion had a strong relationship with learning outcomes. Their exclusion ensures the trend is representative of the majority, but potentially masks important heterogeneity in the enrollment-learning relationship.

Future Research

Item-level psychometric linking: Future work should move beyond aggregate-level regression linking by utilizing Item Response Theory (IRT) on common test items. Rather than relying on the overlap of countries participating in different assessments, researchers could embed common "anchor items" into regional and international assessments. This would allow for a psychometrically valid common scale that does not rely on the assumption that test difficulty ratios remain constant over decades or that the relationship between tests in a few "bridge" countries is representative of an entire region.

Causal identification of growth effects: To address the omitted variable bias in the growth regressions, future research could exploit exogenous shocks to educational quality that are uncorrelated with broader institutional improvements. For example, analyzing the long-term growth impacts of specific, large-scale educational reforms or school construction programs that improved learning outcomes in specific cohorts could help isolate the causal return to cognitive skills from the general effects of state capacity or economic modernization.

Household-based assessment integration: To resolve the selection bias inherent in school-based testing—where marginalized populations are excluded from the data until they enroll—future data collection should integrate household-based learning assessments. By testing a representative sample of the school-age population regardless of enrollment status, researchers could definitively separate genuine learning stagnation from the statistical composition effects caused by the entry of lower-performing students into the school system.

isitcredible.com