February 27, 2026

Editor-in-Chief
AERA Open
American Educational Research Association

Dear Editor,

We are pleased to submit our manuscript, "When Algorithms Mirror Inequality: Structural Encoding of Fairness Disparities in Early Childhood Risk Prediction," for consideration as a regular article in *AERA Open.*

This paper began with a straightforward question: if a school district deploys a machine learning tool to identify at-risk students, will it serve every child equally? The short answer is no—but not for the reasons most people assume.

Using the nationally representative ECLS-K:2011 ($N = 9{,}104$), we evaluated seven algorithms spanning logistic regression to modern gradient boosting. All seven produced nearly identical disparities— Hispanic students flagged at 2.4 times the rate of White students, Black students experiencing nearly four times higher calibration error. That convergence is the central finding: these gaps are not artifacts of any particular model. They are embedded in the data before any algorithm touches it.

Through counterfactual decomposition, we traced the pattern to its source. Statistically aligning Hispanic and White students on baseline characteristics did not merely close the gap—it reversed it. The portion attributable to differences in early cognitive scores and socioeconomic conditions exceeded the observed disparity by 52%, revealing that algorithms were quietly *compressing* inequality rather than amplifying it. Removing SES variables changed almost nothing, because kindergarten test scores already encode the same socioeconomic signal. In effect, these models operate as poverty detectors— and no algorithmic refinement can remedy that from within.

These findings speak directly to *AERA Open*'s readership. First, they reframe the debate from "which algorithm is fairest?" to "what institutional conditions produce data from which fair prediction is even possible?"—a question for educators and policymakers, not just computer scientists. Second, the paper provides a replicable template—combining calibration analysis, intersectional auditing, temporal stability testing, and uncertainty quantification—that researchers can apply to predictive tools increasingly common in K–12 settings.

The manuscript is approximately 10,000 words with 6 figures and 14 tables, plus Supplementary Materials with 9 figures and 8 tables. All analyses use freely available public-use data and reproducible open-source code. This manuscript has not been published or submitted elsewhere, and all authors have approved the submitted version.

We appreciate your time and consideration.

Sincerely,

Kismat Raj Vishwakarma
On behalf of all authors

Corresponding Author
Kismat Raj Vishwakarma
National Institute of Electronics & Information Technology (NIELIT), India
Email: kismat.vishwakarma@gmail.com