




Modeling Parenting Outcomes Using Educational Attainment, Cognitive Stimulation, and Parenting Efficacy with Machine Learning

Tereza. Malá¹, Aditya. Prasetyo^{2*}, Mette. Kristiansen³


¹ Department of Experimental Psychology, Charles University, Prague, Czech Republic

² Department of Educational Psychology, Universitas Gadjah Mada, Yogyakarta, Indonesia



³ Department of Health Psychology, University of Copenhagen, Copenhagen, Denmark

* Corresponding author email address: aditya.prasetyo@ugm.ac.id

Editor

Maximus Monaheng Sefotho¹
Director of the Neurodiversity
Center Department of Educational
Psychology, University of
Johannesburg & University of South
Africa
msefotho@uj.ac.za

Reviewers

Reviewer 1: Mehdi Ghezelsfloo¹
Assistant Professor, Department of Psychology, Gonbadkavos University,
Gonbadkavos, Iran. Email: m.ghezelsfloo@gonbad.ac.ir
Reviewer 2: Mohammadreza Zarbakhsh Bahri¹
Associate Professor, Department of Psychology, Tonekabon Branch, Islamic Azad
University, Tonekabon, Iran. Email: M.Zarbakhsh@Toniau.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Methods section, the statement “a sample of 412 parents from Indonesia selected through stratified random sampling” requires further methodological transparency. Please specify the exact stratification criteria (e.g., region, SES, education level) and proportional allocation procedures, as the current description is insufficient to evaluate sampling validity and representativeness.

In the Introduction, the claim “parenting outcomes constitute a central domain in developmental psychology” is theoretically sound but underdeveloped. You should expand this paragraph by explicitly linking parenting outcomes to established frameworks such as ecological systems theory or transactional models to strengthen the theoretical grounding.

In Table 1 interpretation, the sentence “Cognitive stimulation demonstrated a strong positive relationship with parenting efficacy ($r = 0.52$)” would benefit from a discussion of potential multicollinearity. Please report VIF or tolerance statistics to ensure that predictor overlap does not bias regression estimates.

In Table 2, the regression model is described as explaining “53% of the variance,” but no diagnostic statistics (e.g., residual plots, heteroscedasticity tests) are provided. You should include model diagnostics to support the validity of linear regression assumptions.

In Table 3, the sentence “gradient boosting demonstrated the सर्वोत्तम performance” contains a non-English term (“सर्वोत्तम”), which is inconsistent with academic English writing standards. This should be corrected to maintain linguistic coherence.

Response: Revised and uploaded the new document.

1.2. Reviewer 2

Reviewer:

The paragraph stating “Educational attainment has long been recognized as a foundational determinant of parenting quality” would benefit from a clearer distinction between direct and indirect pathways. You should explicitly model whether education acts as an exogenous predictor or a mediated variable through cognitive stimulation and efficacy, particularly given your later interpretation of indirect effects.

In the cognitive stimulation paragraph, the sentence “It encompasses a range of activities and resources” is overly general. Please operationalize this construct more rigorously by specifying the subscales of the HOME Inventory used and how composite scores were computed in your dataset.

The paragraph introducing parenting efficacy correctly references social cognitive theory, but the sentence “is shaped by environmental factors such as social support” lacks empirical linkage within your model. You should clarify whether these environmental determinants were controlled for or omitted, and discuss the implications of their exclusion.

In the Methods section, the sentence “Data collection was conducted through both online and paper-based formats” raises concerns about mode effects. You should report whether measurement invariance or response bias across modes was tested, as mixed-mode data collection can introduce systematic variance.

The Measures subsection includes “All instruments have demonstrated strong psychometric properties in previous studies,” but does not report specific reliability indices for the current sample. You must provide Cronbach’s alpha (or alternative reliability metrics) for each scale within your sample to establish internal consistency.

In the Data Analysis section, the phrase “handling missing values through multiple imputation” is insufficiently detailed. Please specify the imputation method (e.g., MICE, EM algorithm), number of imputations, and convergence diagnostics to ensure reproducibility.

The statement “normalization of continuous variables” requires clarification regarding the specific transformation used (e.g., z-score standardization, min-max scaling). This is particularly critical given the sensitivity of machine learning algorithms to feature scaling.

In the Findings section, the sentence “Preliminary screening confirmed that all variables met assumptions of normality” is problematic given that several machine learning models (e.g., random forest, gradient boosting) do not require normality assumptions. You should clarify that this assumption pertains only to parametric analyses (e.g., regression).

Response: Revised and uploaded the new document.

2. Revised

Editor’s decision: Accepted.

Editor in Chief’s decision: Accepted.