PUBLISHING INSTITUTE
Journal Website

Article history:
Received 25 January 2026
Revised 05 April 2026
Accepted 18 April 2026
Published Online 10 July 2026

Journal of Adolescent and Youth Psychological Studies

Open peer-review report

E-ISSN: 2981-2526

Hybrid XGBoost–SHAP Prediction of Adolescent Academic Achievement from Cognitive Emotion Regulation, Academic Motivation, Procrastination, and Sleep Quality

Marc. Puigdemont[1], Valentina. Arancibia[2*]

[1] Department of Psychology, University of Andorra, Sant Julià de Lòria, Andorra
[2] Department of Clinical Psychology, Pontifical Catholic University of Chile, Santiago, Chile

*** Corresponding author email address**: varancibia@uc.cl

Editor	Reviewers
Ahmad Amani Associate Professor, Counseling Department, University of Kurdistan, Sanandaj, Iran a.amani@uok.ac.ir	**Reviewer 1:** Seyed Ali Darbani Assistant Professor, Department of Psychology and Counseling, South Tehran Branch, Islamic Azad University, Tehran, Iran. Email: Ali.darbani@iau.ac.ir **Reviewer 2:** Mohammad Salehi Associate Professor, Department of Educational Management, Sari Branch, Islamic Azad University, Sari, Iran. Email: drsalehi@iausari.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Introduction, the paragraph beginning “The interaction of cognitive emotion regulation, motivation, procrastination, and sleep quality reflects a complex network of psychosocial and behavioral determinants” would benefit from a more explicit theoretical framework. The authors should explain why these four domains were selected together and whether their integration is guided by Self-Determination Theory, Self-Regulation Theory, Executive Function models, or another established framework. This would strengthen the conceptual rationale for the predictive model.

The statement “few studies have examined the integrated effects of cognitive emotion regulation, academic motivation, procrastination, and sleep quality using advanced machine learning methodologies” is important but currently lacks a systematic justification. The authors should provide a concise review table or a more detailed synthesis of previous machine-learning studies, specifying sample sizes, algorithms used, outcomes predicted, and methodological limitations to clearly demonstrate the novelty of the present study.

The manuscript states that the model achieved an R^2 of 0.792 in the test set . While impressive, this value is unusually high for psychosocial predictors of academic achievement. The authors should provide additional validation evidence, such as

repeated cross-validation, external validation, nested cross-validation, or bootstrapping, to demonstrate model robustness and reduce concerns regarding optimism bias.

Table 3 reports SHAP importance rankings , including both higher-order constructs (e.g., academic motivation) and lower-level subcomponents (e.g., positive reappraisal, intrinsic motivation, planning strategies). The manuscript should explain how predictors were structured within the model and clarify whether these variables were entered simultaneously. Otherwise, the interpretation of feature importance may be ambiguous due to overlapping construct representation.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The paragraph beginning “In addition to these direct relationships, emerging research suggests that these variables may exert complex indirect and interactive effects” proposes several mediating and moderating mechanisms. However, these mechanisms are not empirically examined in the current study. The authors should clearly distinguish between theoretical propositions and tested hypotheses and avoid implying causal pathways that cannot be evaluated in a cross-sectional design.

In the Methods section, the authors state that participants were recruited through a “multistage cluster sampling procedure” . More methodological detail is required. Specifically, the manuscript should explain how schools were selected, how clusters were defined, how many schools participated, whether proportional allocation was used, and how nonresponse was handled. Without these details, sampling representativeness cannot be adequately evaluated.

The inclusion criterion indicating that students with “documented cognitive impairments or severe psychiatric conditions” were excluded requires clarification. The authors should explain how these conditions were identified, whether school records, parental reports, or clinical documentation were used, and whether ethical approval permitted access to such information.

In the Measures section, descriptions of the CERQ, AMS, PASS, and PSQI are provided , but psychometric properties for the current sample are absent. The manuscript should report internal consistency coefficients (Cronbach’s alpha, McDonald’s omega, or composite reliability) for each scale and, ideally, for relevant subscales within the present dataset.

The sentence “Grade point averages were standardized across participating schools to ensure comparability” requires substantially greater methodological transparency. The authors should describe the standardization procedure in detail, including whether z-score normalization, percentile conversion, or another method was used. This is particularly important because academic achievement serves as the target variable in the predictive model.

In the Data Analysis section, the authors indicate that missing values “representing less than 5% of observations were handled through multiple imputation procedures” . However, the imputation method is not specified. The manuscript should identify the specific imputation algorithm, number of imputations, convergence criteria, and variables included in the imputation model.

The description of hyperparameter tuning states that “learning rate, maximum tree depth, subsampling ratio, minimum child weight, and number of estimators” were optimized . For reproducibility, the authors should provide the complete hyperparameter search space, optimization ranges, final selected values, random seed settings, and software package versions.

Table 1 reports significant correlations among predictors and outcome variables . However, the manuscript does not report confidence intervals for correlation coefficients. Given the large sample size, providing 95% confidence intervals would enhance the precision and interpretability of these findings.

The statement that predictor intercorrelations indicate “the absence of problematic multicollinearity” is not sufficiently supported. The authors should report formal multicollinearity diagnostics such as Variance Inflation Factors (VIF), tolerance values, or condition indices, particularly because several predictors are conceptually related.

Table 2 presents model performance metrics including R^2, RMSE, MAE, and MAPE . However, no benchmark models are provided. The manuscript would be substantially strengthened by comparing XGBoost performance against conventional regression, Random Forest, Support Vector Regression, or other machine-learning algorithms to justify the superiority of the chosen approach.

JAYPS Adolescent and Youth Psychological Studies
E-ISSN: 2981-2526

Authors uploaded the revised manuscript.

2. Revised

Editor's decision after revisions: Accepted.
Editor in Chief's decision: Accepted.

JAYPS Adolescent and Youth Psychological Studies
E-ISSN: 2981-2526