PUBLISHING INSTITUTE
Journal Website

Article history:
Received 17 February 2026
Revised 10 May 2026
Accepted 27 May 2026
Initial Publish 12 June 2026
Published online 01 July 2026

Journal of Assessment and Research in Applied Counseling

Open-Peer Review

E-ISSN: 3041-8518

# Explainable XGBoost Prediction of Academic Impostor Phenomenon from Perfectionistic Concerns, Fear of Failure, and Self-Doubt

Bokros. Vilmos[1,2], Zsuzsa. Csergo[3], Nejc. Rožman[4], Gabor. Harsányi[1,2*]

[1] School Failure Prevention Research Group, Hungarian Academy of Sciences–University of Szeged, 6722 Szeged, Hungary
[2] Institute of Psychology, University of Szeged, 6722 Szeged, Hungary
[3] Doctoral School of Education, University of Szeged, 6722 Szeged, Hungary
[4] Institute of Education, Hungarian University of Agriculture and Life Sciences, 7400 Kaposvár, Hungary

*** Corresponding author email address**: g-harsanyi@edu.u-szeged.hu

| Editor | Reviewers |
|---|---|
| Azizuddin Khan<br>Professor, Psychophysiology Laboratory, Department of Humanities and Social Sciences Indian Institute of Technology Bombay, Maharashtra, India<br>aziz@hss.iitb.ac.in | **Reviewer 1:** Norbert Hetényi<br>Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary.<br>Email: norberthetenyi@ehok.elte.hu<br>**Reviewer 2:** Dilek Soylu<br>Center for Psychotherapy, Ibn Haldun University, Istanbul, Türkiye.<br>Email: dileksoylu@ihu.edu.tr |

## 1. Round 1

### *1.1. Reviewer 1*

Reviewer:

In the Methods, Data Collection Tools section, the description of the academic impostor phenomenon measure mentions the Clance Impostor Phenomenon Scale, but it does not report the number of items, scoring range, response format, interpretation thresholds, or evidence of validity in Hungarian or comparable student populations. The manuscript should include these psychometric details because the outcome variable is central to the entire predictive model. If a translated version was used, translation, back-translation, and cultural adaptation procedures should also be described.

In the Findings, Table 2, the comparison among the baseline model, multiple linear regression, random forest regression, and XGBoost regression is appropriate, but the manuscript should clarify whether the reported RMSE, MAE, and $R^2$ values are based on the independent test set or cross-validation. The table includes both test-like metrics and "cross-validated RMSE," but the distinction is not fully explained in the paragraph. The authors should explicitly state which dataset produced each metric and whether the final performance values were obtained only once on the held-out test set after hyperparameter tuning.

In the Findings, Table 3 and Figure 1, SHAP values are used effectively to interpret model predictions, but the manuscript should provide more methodological detail about SHAP computation. The authors should specify whether TreeSHAP was used, whether SHAP values were calculated on the test set or the entire dataset, and whether the mean absolute SHAP values represent average absolute contributions in the original score metric of the outcome. These details are necessary because SHAP interpretation depends on the data subset and model-output scale used.

Authors revised and uploaded the document.

### *1.2. Reviewer 2*

Reviewer:

In the Methods, Data Collection Tools section, the perfectionistic concerns measure is described generally as "items representing the maladaptive evaluative dimension of perfectionism," but the specific instrument is not named. This is a major methodological limitation because readers cannot evaluate construct validity, scoring, subscales, or comparability with previous studies. The authors should identify the exact scale used, such as the Frost Multidimensional Perfectionism Scale, Almost Perfect Scale-Revised, or another validated instrument, and provide its item number, subscale structure, scoring procedure, and reliability.

In the Methods, Data Collection Tools section, the fear of failure and self-doubt instruments are also described in broad terms without naming the exact scales. The manuscript should provide the full names of the instruments, developers, year, number of items, scoring range, sample items if permitted, and prior evidence of reliability and validity. Since the model's results depend heavily on the measurement quality of these predictors, the current lack of instrument specificity weakens methodological transparency.

In the Methods, Data Analysis paragraph, the manuscript states that the dataset was randomly divided into "80% training and 20% testing," but it should specify whether this split was stratified or whether the outcome distribution was checked across train and test sets. Although stratification is more common in classification, regression studies can still compare outcome means and predictor distributions across splits to ensure that the test set is representative. Reporting this would strengthen confidence that the model evaluation was not influenced by an unbalanced random partition.

In the Methods, Data Analysis paragraph, the authors list hyperparameters such as "learning rate, maximum tree depth, number of estimators, subsampling ratio, column sampling ratio, minimum child weight, gamma, and regularization parameters," but the final selected values are not reported in the findings. For reproducibility, the manuscript should include the optimized hyperparameter values, the cross-validation strategy, the tuning method, the number of folds, the search space, and the evaluation metric used during tuning. This is especially important in machine learning studies because model performance can vary substantially depending on tuning decisions.

In the Findings, Table 1, the reported correlations are useful, but the manuscript should also report confidence intervals for the correlation coefficients. For example, the correlation between self-doubt and academic impostor phenomenon is reported as $r = .66$, $p < .001$, but a confidence interval would provide more information about the precision of the estimate. Since the sample size is relatively adequate, reporting confidence intervals would improve statistical transparency and align the results section with contemporary reporting standards.

Authors revised and uploaded the document.

## 2. Revised

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



JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518