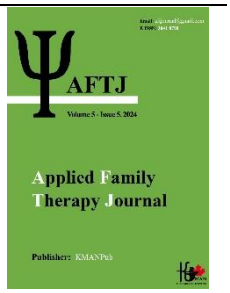
KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 13 May 2026
Revised 16 August 2026
Accepted 24 August 2026
Published online 01 September 2026

Applied Family Therapy Journal

OPEN PEER-REVIEW REPORT



E-ISSN: 3041-8798

Predicting Co-Parenting Quality From Marital Adjustment, Parenting Self-Efficacy, and Psychological Distress Using Gradient Boosting Regression

Pieter. Verhaegen[1], Mariana. Coutinho[2*], Tomas Jankauskas[3]

[1] Department of Social Psychology, Ghent University, Ghent, Belgium
[2] Department of Clinical Psychology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[3] Department of Social Psychology, Vilnius University, Vilnius, Lithuania

*** Corresponding author email address**: mariana.coutinho@ufrj.br

Editor	Reviewers
Habib Hadianfard Affiliation: Professor, Department of Psychology, Shiraz University, Iran hadianfd@shirazu.ac.ir	**Reviewer 1:** Mohammadreza Zarbakhsh Bahri Associate Professor ،Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran. Email: M.Zarbakhsh@Toniau.ac.ir **Reviewer 2:** Farideh Dokanehi Fard Associate Professor, Counseling Department, Roudehen Branch, Islamic Azad University, Roudehen, Iran. Email: f.dokaneifard@riau.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

The paragraph beginning “Psychological distress represents a third major factor that may undermine effective co-parenting” should more clearly justify the use of a nonspecific distress construct rather than separate depression, anxiety, stress, or parenting-specific distress measures. The K10 captures general psychological distress but does not identify whether depressive symptoms, anxiety symptoms, parenting stress, or other emotional difficulties drive prediction. This distinction is particularly important because much of the literature cited in this section concerns parenting stress, postpartum depression, or anxiety rather than nonspecific distress. The authors should either strengthen the conceptual rationale for using a global distress index or explicitly acknowledge that the K10 represents a broad transdiagnostic indicator whose predictive role cannot be decomposed into specific symptom domains.

The Introduction paragraph asserting that “Gradient Boosting Regression is particularly suitable for this purpose because it constructs an ensemble of sequential decision trees” is currently too general for a manuscript positioning machine learning as a methodological contribution. The authors should explain why Gradient Boosting was selected over alternative nonlinear algorithms such as Random Forest, XGBoost, CatBoost, support vector regression, or neural networks. The rationale should

relate to the sample size, number of predictors, expected nonlinearities, tabular nature of the data, and interpretability requirements. Without such justification, the choice of algorithm appears somewhat arbitrary, particularly because only three predictors were included and nonlinear effects could alternatively be modeled using restricted cubic splines or generalized additive models.

The description of the train-test procedure states that "The data set was randomly divided into an 80% training set and a 20% holdout test set using a fixed random seed." A single random split can generate unstable estimates, particularly with a test sample of only 93 participants. The authors should either justify this approach or consider nested repeated cross-validation, bootstrap optimism correction, or repeated train-test evaluation to quantify uncertainty. At minimum, 95% confidence intervals should be reported for test-set R^2, RMSE, and MAE. Without uncertainty estimates, the observed improvement from $R^2 = .49$ for linear regression to $R^2 = .61$ for Gradient Boosting may appear more definitive than warranted.

Table 2 reports only five "candidate models," including a final model with 300 estimators, learning rate .03, maximum depth 2, minimum leaf size 8, and subsample .80. The manuscript should describe the full hyperparameter search space and tuning strategy rather than presenting a small selection of apparently hand-picked alternatives. It is unclear whether grid search, randomized search, Bayesian optimization, or manual tuning was used. The exact ranges and values evaluated for each hyperparameter should be reported. Additionally, the authors should state whether early stopping was considered and how stochastic subsampling interacted with random-state specification to ensure reproducibility.

Table 3 concludes that "Gradient Boosting Regression" performed better than linear regression because test R^2 increased from .49 to .61 and RMSE decreased from 18.23 to 15.95. This comparison requires stronger statistical evaluation. The authors should quantify whether the difference in predictive errors is robust using bootstrap confidence intervals or a paired resampling procedure applied to individual test-set prediction errors. Reporting only point estimates does not establish that the improvement is statistically or practically meaningful. The manuscript should also consider reporting normalized RMSE or comparing RMSE with the standard deviation of the outcome to help readers understand the magnitude of prediction error relative to natural variation in co-parenting quality.

Table 4 combines "Normalized Feature Importance," "Permutation Importance in R^2," and "Mean Absolute SHAP Value," which is commendable, but the manuscript needs to distinguish the interpretations of these metrics more rigorously. Impurity-based feature importance can be biased, permutation importance measures dependence of predictive performance on a feature, and mean absolute SHAP values quantify average contribution magnitude rather than causal influence. The phrase "marital adjustment was the most influential predictor" should therefore be replaced with more precise wording such as "marital adjustment showed the highest predictive importance." The authors should also report the stability of these rankings across resamples because feature importance can fluctuate substantially, particularly when predictors are correlated.

Response: Revised and uploaded the manuscript.

1.2. Reviewer 2

Reviewer:

In the Study Design and Participants section, the sentence "The final analytic sample therefore consisted of 462 parents" requires a formal justification of whether this sample size is adequate for the proposed machine-learning analysis. Conventional power analysis for multiple regression is insufficient because the study emphasizes predictive performance. The authors should report either a learning-curve analysis, simulation-based precision assessment, or rationale based on the expected stability of cross-validation and holdout metrics. With only three predictors, 462 cases may be sufficient, but the manuscript should provide quantitative justification rather than assuming adequacy. The reported 80/20 split also leaves only 93 participants in the test set, which may produce substantial uncertainty in R^2 and error estimates.

The participant section states that respondents were recruited "from community settings, family health services, schools, parenting networks, and online parent groups in several Brazilian states," but the recruitment process is insufficiently described for reproducibility and assessment of selection bias. The authors should specify recruitment dates, the proportions recruited

AFTJ Applied Family Therapy Journal
E-ISSN: 3041-8798

through each channel, whether advertisements were targeted, whether both members of the same couple could participate, and whether duplicate household participation was detected. If both partners from the same family entered the dataset, the assumption of independent observations would be violated. This issue is especially important in co-parenting research because responses from partners belonging to the same dyad are inherently dependent.

In the eligibility criteria, the manuscript specifies that participants had "at least one child between 3 and 12 years of age living in the household." The theoretical or methodological reason for restricting the child's age to this range is not explained. Co-parenting dynamics vary substantially across developmental stages, and parenting responsibilities for preschool-aged children may differ from those for preadolescents. The authors should justify this age restriction and report the age distribution of the target child. If participants had multiple eligible children, the procedure used to determine which child they considered when completing the questionnaires must also be specified, because ambiguity regarding the reference child can introduce measurement noise.

The Data Collection Tools section reports that "the Brazilian Portuguese version was administered using standardized translation and adaptation procedures" for the Coparenting Relationship Scale. This wording is inadequate unless a previously validated Brazilian version was actually used. The manuscript should provide the source of the Brazilian adaptation, describe its psychometric validation, and cite the adaptation study. If the authors translated the instrument specifically for this research, they should report forward translation, back-translation, expert review, pilot testing, and cultural adaptation procedures rather than simply stating that standardized procedures were used. Similar clarification is necessary for the Portuguese versions of the Revised Dyadic Adjustment Scale, Parenting Sense of Competence Scale, and K10.

The manuscript states that the Coparenting Relationship Scale has "possible scores ranging from 0 to 210." This scoring specification should be verified carefully against the exact response formats and scoring instructions of the version used. The original CRS includes items with differing response formats, and straightforward summation may not necessarily produce the stated range unless items were transformed. The authors should describe exactly how individual items were scored, how reverse-coded items were handled, whether dimension scores were standardized before aggregation, and how the total scale range was derived. This is essential because the dependent variable forms the basis for every subsequent predictive metric.

In the Data Analysis section, the sentence "Missing data accounted for less than 1.5% of item-level responses, and missing values remaining after scale scoring were imputed using the median calculated within the training data to prevent information leakage" is methodologically appropriate in principle but insufficiently transparent. The authors should report the number of participants and variables affected, whether missingness patterns were examined, and whether data were plausibly missing completely at random or missing at random. More importantly, if scale-level scores could be computed when only some items were missing, the manuscript should specify the minimum number of completed items required. Median imputation should also be performed separately within each training fold during cross-validation rather than once on the entire training set if preprocessing is included in the tuning workflow.

Response: Revised and uploaded the manuscript.

2. Revised

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

AFTJ Applied Family Therapy Journal
E-ISSN: 3041-8798