Journal Website

Article history:
Received 24 October 2025
Revised 03 February 2026
Accepted 10 February 2026
Published online 10 April 2026

Journal of Adolescent and Youth Psychological Studies

Open peer-review report

E-ISSN: 2981-2526

Predicting Self-Esteem Instability Using Rejection Sensitivity, Daily Stress Reactivity, Social Feedback Valence, and Emotional Reactivity in LSTM Models

Takumi. Shibata[1], Álvaro. Montoya[2*]

[1] Department of Cognitive Psychology, Waseda University, Tokyo, Japan
[2] Department of Social Psychology, Autonomous University of Madrid, Madrid, Spain

*** Corresponding author email address**: alvaro.montoya@uam.es

Editor	Reviewers
John S. Carlson Distinguished Professor of the Department of Educational Psychology, Michigan State University, East Lansing, MI, United carlsoj@msu.edu	**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:** Kamdin Parsakia Department of Psychology and Counseling, KMAN Research Institute, Richmond Hill, Ontario, Canada. Email: kamdinarsakia@kmanresce.ca

1. Round 1

1.1. Reviewer 1

Reviewer:

The description of the EMA design and measurement strategy is generally clear, but key methodological details are either dispersed or underspecified: the sampling schedule (exact number of prompts per day, timing scheme, compliance thresholds), inclusion/exclusion criteria, and handling of non-adherence (e.g., minimum number of observations per participant to be retained) should be presented more explicitly and compactly, ideally in a dedicated subsection or table to allow readers to fully evaluate ecological validity and potential selection biases.

The operationalization of self-esteem instability is central to the study, yet the metric used (e.g., within-person SD, RMSSD, or another time-series variability index) and the exact computation pipeline are not described with sufficient technical precision; given that different instability indices capture distinct statistical and temporal properties, the authors should (a) formally define the chosen index with a formula, (b) justify its selection based on prior work, and © clarify whether instability is computed per time window, per day, or across the entire series and how this aligns with the prediction targets of the LSTM.

The proposed clinical and technological implications, especially the vision of JITAI-type smartphone interventions based on real-time predictions of self-esteem drops, are intriguing and forward-looking, but they currently rest on a single observational dataset and offline modeling; the manuscript would benefit from a clearer delineation between what has been empirically demonstrated (offline predictive performance in a specific sample) and what remains speculative (real-time deployment, user acceptability, ethical safeguards), perhaps by outlining concrete next steps for feasibility studies, human-in-the-loop evaluations, or pilot intervention trials.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The modeling section provides a strong starting point, indicating the use of LSTM with recurrent dropout, Adam optimizer, and rolling-origin cross-validation; nevertheless, the reproducibility of the machine-learning pipeline is limited by missing hyperparameter details (e.g., number of layers and units, activation functions, look-back window length, batch size, number of epochs, early stopping criteria, regularization coefficients), and the authors are encouraged to either include a full hyperparameter table in the main text or provide code/parameter settings in a public repository to meet current standards for transparent ML research.

The comparative evaluation between LSTM, multiple linear regression, and standard RNN is an important contribution, but the presentation of results focuses primarily on point estimates of RMSE/MAE without sufficiently addressing the statistical uncertainty and practical significance of these differences; the authors should consider adding confidence intervals or bootstrap distributions for performance metrics, conduct formal tests of differences where appropriate, and discuss effect sizes in terms of real-world interpretability (e.g., what does a given reduction in RMSE mean in units of self-esteem and in terms of clinical or practical relevance).

The treatment of feature importance via permutation importance is a welcome attempt to open the "black box" of the LSTM, yet the current exposition remains rather high-level; the paper would be strengthened by (a) clarifying how permutation importance was implemented in a time-series context (e.g., whether temporal structure was preserved during permutations), (b) reporting uncertainty around importance estimates, and © complementing permutation scores with additional interpretability tools (e.g., partial dependence plots, time-resolved importance analyses, or attention-like visualizations) to deepen mechanistic insights into how emotional reactivity and rejection sensitivity jointly shape predicted instability.

The discussion effectively links the findings to broader theories of attachment, rejection sensitivity, and dynamic conceptualizations of self-esteem, but it sometimes moves quickly from predictive associations to quasi-causal language; the authors are encouraged to more clearly separate what their design and analyses can and cannot support in terms of causal inference, explicitly acknowledging that even high predictive accuracy from LSTM does not by itself establish causal pathways, and to phrase clinical implications in a more cautiously conditional manner.

The limitations section correctly notes key issues such as self-report EMA, sample homogeneity, and the black-box nature of deep learning models, yet it could engage more critically with two additional constraints: first, the possibility of overfitting despite the use of dropout and rolling-origin CV (e.g., by briefly reporting training vs. validation learning curves or sensitivity analyses); and second, the potential impact of missing-data imputation strategies (forward filling and moving averages) on temporal dynamics, including a discussion of how alternative imputation methods might alter estimates of instability and predictor relationships.

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