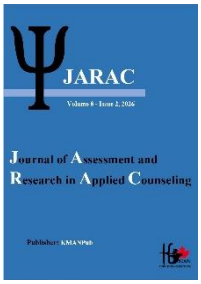
KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 23 December 2025
Revised 11 March 2026
Accepted 17 March 2026
Published online 01 April 2026

Journal of Assessment and Research in Applied Counseling

Open-Peer Review



E-ISSN: 3041-8518

Predicting Therapist Effectiveness by Empathy Accuracy, Session Synchrony, Linguistic Alignment, and Reflective Depth: A Machine Learning Analysis

Anastasios. Kyrillidis[1*], Jacqueline. Woolley[1], Julie. Larson[2]

[1] Department of Psychology, University of Guelf, Guelf, Canada
[2] Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

*** Corresponding author email address**: anakyrillidis@uoguelph.ca

Editor	Reviewers
Valiollah Farzad Associate Professor, Department of Psychology and Counseling, KMAN Research Institute, Richmond Hill, Ontario, v.farzad@kmanresce.ca	**Reviewer 1:** María José Ibañez Study Group on Advances in Psychological Measurement, National University of San Marcos, Lima, Peru. Email: jose.ibnez@upn.pe **Reviewer 2:** Matthew McNally Department of Clinical Health Psychology, University of Manitoba, Winnipeg, MB, Canada. Email: mattemcnally21@gmail.com

1. Round 1

1.1. Reviewer 1

Reviewer:

The phrase “therapists were recruited through professional counseling associations and clinical networks” suggests convenience sampling. Please clarify the sampling strategy and discuss potential selection bias, especially regarding therapists who are more technologically engaged or research-inclined.

In the Measures section, the description “Empathy accuracy was assessed using a structured empathic inference task…” is insufficiently detailed. Please specify the psychometric properties (e.g., reliability, validity indices) of this task, and whether it has been previously validated in clinical populations.

The operationalization of “session synchrony…via behavioral and physiological indicators” lacks clarity regarding temporal resolution. Please specify whether synchrony was computed at the second-by-second level, windowed intervals, or aggregated across sessions, as this directly impacts interpretability of the findings.

In Table 3 (Feature Importance), SHAP values are reported as means, but variability (e.g., standard deviation or distribution) is not provided. Including dispersion metrics would improve interpretability and robustness assessment.

There is a duplication error: two tables are labeled as “Table 3.” This is a structural inconsistency that must be corrected to maintain academic clarity and referencing integrity.

In the Interaction Effects section, the phrase “interaction strength” is not formally defined. Please clarify whether this refers to SHAP interaction values, partial dependence interactions, or another metric.

Authors revised and uploaded the document.

1.2. Reviewer 2

Reviewer:

In the linguistic alignment measurement, the sentence “using embedding-based models” is overly vague. Please specify which embedding models (e.g., BERT, Word2Vec) were used, including language model versioning and training corpus, to ensure reproducibility.

In the Reflective Depth measurement, the phrase “a validated coding system” requires citation and elaboration. Please describe coder training, inter-rater reliability (e.g., ICC or Cohen’s kappa), and how discrepancies were resolved.

In the Data Analysis section, the statement “handling of missing data through multiple imputation” is insufficiently specified. Please indicate the imputation method (e.g., MICE), number of imputations, and whether missingness mechanisms (MCAR, MAR, MNAR) were tested.

The sentence “feature engineering to extract temporal and interaction-based metrics” lacks transparency. Please provide examples of engineered features and justify their theoretical relevance to psychotherapy processes.

In the modeling section, while multiple algorithms are mentioned, there is no discussion of hyperparameter tuning. Please specify optimization procedures (e.g., grid search, Bayesian optimization) and evaluation protocols to ensure comparability across models.

In Table 1, although correlations are reported, there is no indication of statistical significance levels or confidence intervals. Please include p-values or CI ranges to support interpretation of the associations.

In the paragraph following Table 1, the claim “without severe multicollinearity concerns” should be supported by formal diagnostics (e.g., VIF values). Please report these indices explicitly.

In Table 2, model performance is reported, but no baseline model (e.g., linear regression) is included. Including a baseline would strengthen the argument that machine learning provides incremental predictive value.

The sentence “deep neural network achieved the highest predictive accuracy…$R^2 = 0.79$” requires clarification of whether this is test-set performance or cross-validated performance. Please ensure strict separation of training and evaluation metrics to avoid optimistic bias.

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