

# Identifying Digital Behavior Profiles via Usage Patterns, Reward Sensitivity, and Social Reinforcement with Machine Learning Analysis

Mateusz. Kaczmarek<sup>1</sup>, Emily. Cartwright<sup>2\*</sup>, Gabriela. Uzcátegui<sup>3</sup>

<sup>1</sup> Department of Psychology, University of Warsaw, Warsaw, Poland

<sup>2</sup> Department of Clinical Psychology, University of Toronto, Toronto, Canada

<sup>3</sup> Department of Social Psychology, Central University of Venezuela, Caracas, Venezuela

\* Corresponding author email address: emily.cartwright@utoronto.ca

## Editor

Sergii Boltivets  
Chief Researcher of the Department of Scientific Support of Social Formation of Youth. Mykhailo Drahomanov University, Ukraine  
sboltivets@ukr.net

## Reviewers

**Reviewer 1:** Mehdi Ghezelsfloo  
Assistant Professor, Department of Psychology, Gonbadkavos University, Gonbadkavos, Iran. Email: m.ghezelsfloo@gonbad.ac.ir  
**Reviewer 2:** Abotaleb Saadati Shamir  
Assistant Professor, Department of Educational Sciences, University of Science and Research, Tehran, Iran. Email: psychology@iaau.ac.ir

## 1. Round 1

### 1.1. Reviewer 1

Reviewer:

In the paragraph discussing reinforcement mechanisms, the sentence “Social media platforms, for instance, are deliberately designed to provide intermittent rewards” would benefit from more precise terminology (e.g., variable ratio reinforcement schedules) and citation alignment with behavioral learning theory to avoid oversimplification of reinforcement paradigms .

The section introducing reward sensitivity states “Individuals with higher reward sensitivity tend to exhibit stronger motivational responses to positive stimuli,” yet the manuscript does not clarify whether BAS subscales were treated as separate predictors or aggregated; this ambiguity undermines interpretability of the clustering inputs and should be explicitly resolved .

In the paragraph on social reinforcement, the phrase “digital reinforcement is immediate, quantifiable, and socially visible” is theoretically sound but requires empirical grounding; the authors should specify how these dimensions were operationalized in their adapted scale and whether factor analysis confirmed these dimensions in the current sample .

The statement “Persuasive design techniques...are engineered to maximize user engagement” would benefit from a more critical methodological linkage to the current study, as no platform-level variables were included in the model; this creates a conceptual-methodological disconnect that should either be addressed or explicitly acknowledged .

The classification results stating “random forest model demonstrating superior accuracy” should include confidence intervals or cross-validation strategy (e.g., k-fold CV) to ensure that reported performance is not due to overfitting or sample-specific variance .

In the PCA visualization description, the sentence “first two principal components accounted for a substantial proportion of variance” is vague; the exact percentage of explained variance must be reported to support claims of effective dimensionality reduction .

Response: Revised and uploaded the new document.

### 1.2. Reviewer 2

Reviewer:

In the Methods section, the sentence “Data were collected using standardized self-report instruments” is insufficiently precise; the authors should report reliability indices (Cronbach’s alpha) for the current sample rather than relying solely on prior literature, as internal consistency is sample-dependent .

The description “a pilot test with 30 participants was conducted” lacks methodological detail; it is unclear whether pilot data were included in the final analysis, how modifications were made based on pilot feedback, and whether reliability estimates were recalculated post-adaptation .

In the Data Analysis section, the phrase “handling of missing values using multiple imputation” requires elaboration on the imputation model (e.g., MICE, number of imputations, convergence diagnostics), as these parameters critically influence downstream clustering results .

The sentence “K-means clustering and hierarchical clustering methods were applied” does not specify distance metrics (e.g., Euclidean, Ward linkage), initialization strategy, or convergence criteria, which are essential for reproducibility and evaluation of clustering robustness .

In reporting cluster validation, “optimal cluster number determined using the silhouette coefficient and the Davies–Bouldin index” is appropriate, but the manuscript does not provide the actual index values or justification for selecting the three-cluster solution over alternatives (e.g., k=2 or k=4), which weakens methodological transparency .

The statement “cluster stability was validated באמצעות bootstrapping procedures” contains a language inconsistency (“באמצעות”), which must be corrected, and additionally requires clarification on the number of bootstrap resamples and stability metrics used (e.g., Jaccard similarity) .

In Table 1 interpretation, the sentence “moderate to high levels across most behavioral and psychological constructs” is descriptive but lacks statistical justification; the authors should define thresholds or compare against normative benchmarks to substantiate this classification .

In the Results section, the phrase “clear separation of centroids across clusters indicates robust differentiation” is interpretative but insufficient without reporting inter-cluster distance metrics or ANOVA comparisons across clusters to statistically validate these differences .

Response: Revised and uploaded the new document.

## 2. Revised

Editor’s decision: Accepted.

Editor in Chief’s decision: Accepted.