Preregistration

Many-analysts Psychedelics Project: Underlying Factors of Psychedelic Questionaires

Parsa Yousefi[1], Agnese Ferrante[1], Julian Karch[1]

[1] Leiden University

02. April 2024

Study Information

Title	Many-analysts Psychedelics Project: Underlying Factors of Psychedelic Questionaires
Description	This project, focusing on the analysis of psychedelic experiences, addresses a gap in psychedelic research: the diverse and sometimes conflicting scales used to measure such experiences. Despite the proliferation of research scales, there is a lack of consensus regarding the constructs being measured and their underlying factor structures. Our objective is to assess the factor structure of the most commonly employed scales and their potential overlap, utilizing factor and latent class analysis.
	The project revolves around two primary questions:

1. Whether the current dataset supports a common factor structure ("mystical experience") underlying the diverse scales measuring the psychedelic experience.

2. Identification of the best predictors of well-being according to our analysis.

- Which qeustionaires?

Hypotheses   Enter your response here.

Design Plan

Study type   **Observational Study**. Data will be analyzed from existing scales that measure psychedelic experiences, without the need for direct intervention or random assignment of treatments. This involves the statistical examination of previously collected datasets.

- A many analysts approach will be employed, with multiple teams analyzing the same dataset to explore the same reserach qeustion regarding the factor structure of the scales.

Blinding   No blinding is involved in this study, as it is based on the analysis of existing datasets without direct interaction with subjects. The analysis focuses on uncovering patterns within the data using statistical methods, independent of any experimental manipulation or treatment assignment.

Study design   Our study employs a many-analyst, observational design to explore factor structures in psychedelic research scales within a newly collected dataset. This approach leverages diverse statistical analyses, without specific counterbalancing, to assess commonalities and distinctions across scales.

Randomization   Randomization is directly not applicable to the design of our study, as it involves the analysis of a dataset that has already been collected by collaborators rather than the direct allocation of treatments or interventions to subjects. We will split

the dataset across participants into a training and a test set. This splitting will be done randomly, ensuring that the two sets are independent of each other. The training set will be used to identify the factor structure of the scales, while the test set will be used to validate the model.

Sampling Plan	Our study is part of a many-analysts project and will analyze a newly collected dataset. Given the collaborative nature of this project, our analysis represents one of many, contributing to a broader understanding through the synthesis of findings across different analytical teams. At this stage, the exact number of samples in the dataset is not determined, as we do not yet have access to the data.
Existing data	**Registration prior to accessing the data**. As of the date of submission, the data exist, but have not been accessed by our team.
Explanation of existing data	Since our project centers around a newly collected dataset for a many-analyst approach, the concept of "existing data" doesn't directly apply as traditionally defined. However, to ensure the integrity and impartiality of our analysis, we have established protocols to prevent any one analyst or team from having advance access to the dataset or its summary statistics before the analysis begins. The dataset will be revealed to all participating analysts simultaneously, after the completion of data collection, to guarantee that all analyses are conducted without prior knowledge of the data's characteristics.
Data collection procedures	For detailed information on the data collection process, including participant recruitment, selection criteria, and survey methodology, please refer to the original study documentation. As this project involves a many-analyst approach to analyze the dataset, specific details on data collection are beyond the scope of our analysis plan. For comprehensive insights into the procedures and methodologies employed in gathering the data, we recommend consulting the primary source of the dataset.
Sample size	As this project employs a many-analyst approach, the exact sample size to be analyzed is not determined by out team.

Sample size rationale	Given the collaborative nature of our project and the diversity of statistical methods to be employed by various analysts, our primary concern is to collect as rich and varied a dataset as possible. The sample size, therefore, will be influenced by practical considerations such as the availability of participants and the resources available for data collection. The goal is to maximize the sample size within these constraints.
Stopping rule	A stopping rule does not apply to our project as we are not directly involved in the data collection process.

Variables

Manipulated variables	Not applicable. This observational study does not involve experimental manipulation of variables. Our analyses are aimed at understanding patterns within an existing dataset.
Measured variables	For Research Question 1 (common factor structure):
	Measured Variables: The sum scores of the scales MEQ, CEQ, AWES, ESAT, EDI, SWLS, NADA_S, ASC11D, ASC11DShort, EISI, INOE, APEQ_S, LAP, and APEI. These sum scores are calculated based on responses to individual items within each scale.
	For Research Question 2 (best predictor of well-being):
	Outcome Variable: Well-being, operationalized through scale scores related to psychological well-being (Satisfaction with Life Scale; SWLS). Predictors: Sum scores of all analyzed scales,
	For Research Question 3 (most important conclusion/finding):
	Predictors: Additional demographic or psychometric data collected as part of the dataset, such as age, gender, prior psychedelic use, and psychological traits or states before the psychedelic experience

Indices
    we will focus on a subset of indices that offer a balanced view of our models' fit to the data. Our selection rationale aligns with current best practices in structural equation modeling (SEM) and factor analysis:

Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI): As incremental fit indices, both CFI and TLI will help us assess the relative improvement of our model fit compared to a baseline (null) model. We aim for values above .95, indicating a good fit to the data.

Root Mean Square Error of Approximation (RMSEA): As an absolute fit index, RMSEA provides a measure of how well the model, with unknown but optimally chosen parameter estimates, fits the population covariance matrix. Values less than .05 indicate a close fit, and values up to .08 represent a reasonable error of approximation in the population.

Standardized Root Mean Square Residual (SRMR): This index assesses the standardized difference between the observed and predicted correlations, offering a direct measure of the average discrepancy per correlation. Values less than .08 are indicative of a good fit.

Sample Size Considerations for Fit Indices: We acknowledge the sensitivity of chi-square and RMSEA to sample size, with larger samples often leading to significant chi-square values and potentially inflated RMSEA values. As such, we will interpret these indices in the context of our sample size, adhering to recommendations for adjusting expectations of fit indices based on sample characteristics.

Analysis Plan

Statistical models
    Confirmatory Factor Analysis (CFA): To explore the common factor structure across psychedelic experience scales, we will use CFA. This involves specifying a model where each scale's sum score is treated as an indicator of a latent factor (e.g., a "Mystic" factor) that we hypothesize to underlie these measures.

Lasso Regression: To identify the best predictors of well-being, we will use Lasso regression, which is particularly suited for models that might suffer from multicollinearity or where the number of predictors is large. This method helps in variable selection by penalizing the absolute size of the regression coefficients. The

outcome variable will be well-being, measured by scales such as the Satisfaction With Life Scale (SWLS), with predictors including sum scores from the included scales.

Model Comparison: Nested models, from more restrictive (a single-factor model for each questionnaire) to less restrictive (hierarchical and bi-factor models), will be compared using ANOVAs to understand which model best fits the data.

Transformations   Data will be checked for normality, and transformations will be applied as necessary.

Inference criteria   Inferences will be based on:

p-values, with a standard criterion of p < .05 for determining statistical significance. Confidence intervals for parameter estimates to assess the precision of our estimates. Fit indices for CFA, with cut-offs of CFI and TLI > .95, RMSEA < .05 (good fit) to .08 (acceptable fit), and SRMR < .08 indicating good model fit.

Data exclusion   Not applicaple, as our team will receive a cleaned dataset from the project coordinators.

Missing data   Not applicaple, as our team will receive a cleaned dataset from the project coordinators.

Exploratory analyses (optional)   Exploratory factor analyses (EFA) may be conducted to investigate the dimensionality of the psychedelic experience scales, especially in the early stages of the analysis to inform the CFA models. Additionally, exploratory analyses may be performed to investigate demographic or psychological characteristics that could moderate the relationship between psychedelic experiences and well-being.

Other

Other (Optional)   Enter your response here.

References