

Study Protocol and Statistical Analysis Plan

Official Title

Development and validation of the Excessive Emotional Behaviors Scale:
long and short forms among Chinese adults with emotional distress

Short Title

Excessive Emotional Behaviors Scale Development

Document Date

23 April 2026

1. Background

Emotional disorders such as anxiety and depression represent a major contributor to the global burden of disease due to their high prevalence, recurrent course, and substantial impact on social and occupational functioning. In recent years, the transdiagnostic approach has emerged as a dominant paradigm in the study and treatment of emotional disorders. This approach is based on the assumption that different emotional disorders share common underlying psychological mechanisms.

Within this framework, emotional behaviors are conceptualized as a core mechanism involved in the onset, maintenance, and exacerbation of emotional disorders. Emotional behaviors refer to behavioral responses triggered by intense emotional experiences, typically aimed at rapidly reducing or controlling emotional distress. Although these behaviors may provide short-term relief, they often contribute to long-term maintenance of emotional problems and functional impairment.

The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders identifies emotional behaviors as a key treatment target, emphasizing the reduction of avoidance and maladaptive emotional action tendencies. In parallel, within the context of Chinese clinical practice, Mindfulness Intervention for Emotional Distress has been proposed, alongside the psychopathological diamond model of emotional distress. This model suggests that emotional distress is characterized by systematic dysfunction across four domains: life engagement, distress tolerance, emotional behaviors, and cognitive flexibility. Among these, excessive emotional behaviors are considered a central hub linking emotional experiences to functional impairment.

Despite their theoretical importance, there is currently no specialized instrument designed to directly assess excessive emotional behaviors. Existing measures, such as those assessing experiential avoidance, primarily focus on internal experiences or cognitive attitudes rather than observable behavioral responses. Furthermore, many existing instruments are lengthy and contain overlapping constructs, limiting their practical utility.

Therefore, it is necessary to develop a concise, theoretically grounded, and psychometrically sound scale to assess excessive emotional behaviors.

2. Objectives and Hypotheses

2.1 Overall Objective

The primary aim of this study is to develop and validate a self-report scale measuring excessive emotional behaviors, grounded in the Unified Protocol and Mindfulness Intervention for Emotional Distress frameworks.

2.2 Specific Objectives

To define the construct and dimensions of excessive emotional behaviors.

To identify real-world behavioral patterns through qualitative methods.

To examine the factor structure, reliability, and validity of the scale.

To explore associations between excessive emotional behaviors and emotional symptoms.

2.3 Hypotheses

Excessive emotional behaviors represent a multidimensional construct, including:

behavioral avoidance

cognitive avoidance

safety behaviors

emotion-driven actions

The scale will demonstrate good internal consistency and structural validity.

Scale scores will be positively associated with anxiety and depressive symptoms.

Excessive emotional behaviors will be related to, but distinguishable from, experiential avoidance.

3. Study Design

This study adopts a mixed-methods design consisting of:

Study 1: Qualitative study (cognitive interviews)

Study 2: Quantitative study (scale development and validation)

4. Study 1: Qualitative Phase

4.1 Objective

To evaluate item clarity, relevance, and ecological validity through cognitive interviews.

4.2 Participants

Approximately 18 participants with anxiety or depressive distress will be recruited.

Inclusion Criteria

Age ≥ 18 years

Experience of emotional distress

Ability to understand and communicate in Chinese

Provision of informed consent

Exclusion Criteria

Acute psychosis or severe psychiatric disorders

Cognitive impairment

4.3 Procedure

Semi-structured interviews (~50 minutes) will include:

Rapport building

Cognitive interviewing (think-aloud + probing)

Structural evaluation

4.4 Data Analysis

Thematic analysis will be conducted:

Open coding

Theme development

Mapping to item pool

Two independent coders will analyze data.

5. Study 2: Quantitative Phase

5.1 Participants

Participants will be recruited using random sampling.

Sample size:

5–10 participants per item, 550.

5.2 Measures

Excessive Emotional Behaviors Scale (Preliminary Version)

Developed based on qualitative phase.

Experiential Avoidance Measures

Acceptance and Action Questionnaire–II

Brief Experiential Avoidance Questionnaire

Anxiety Measure

Generalized Anxiety Disorder Scale–7

Depression Measure

Patient Health Questionnaire–9

6. Statistical Analysis Plan

This study aims to develop the Excessive Emotional Behavior Scale to assess emotional behavior patterns exhibited by individuals experiencing emotional distress. Based on the theoretical framework, the initial item pool will consist of 40 items covering four hypothesized dimensions: behavioral avoidance, cognitive avoidance, safety behavior, and emotion-driven behavior. To ensure the stability of the scale structure and the psychometric quality of the instrument, a two-stage independent-sample design will be used for scale development and validation. In the first stage, an independent sample will be recruited for item analysis and exploratory factor analysis to establish the preliminary structure of the scale. In the second stage, a new independent sample will be recruited for confirmatory factor analysis, as well as reliability and validity testing, to determine the final item composition and dimensional structure of the formal scale.

6.1 Sample Use Strategy

A two-stage independent-sample design will be adopted in this study. In the first stage, one independent sample will be recruited for item analysis and exploratory factor analysis to preliminarily screen items and explore the latent factor structure of the scale. In the second stage, another new independent sample will be recruited for confirmatory factor analysis to test the stability and replicability of the factor structure identified in the first stage. This design helps avoid the risk of overfitting that may arise when structural exploration and structural validation are conducted within the same sample, thereby improving the rigor of structural validation and the

external stability of the findings.

The first-stage sample will be used only for item analysis and exploratory factor analysis and will not be used for confirmation of the final structural model. The second-stage sample will be used independently for confirmatory factor analysis, internal consistency testing, and the evaluation of structural validity, convergent validity, discriminant validity, and criterion-related validity. If the sample size in the second stage permits, additional competing model comparisons and supplementary model testing will also be conducted.

6.2 Data Preparation and Preprocessing

Before formal analyses, data cleaning and preprocessing will be conducted. This will include checking for missing values, abnormal response patterns, extreme values, duplicate responses, and unusually short completion times. Clearly invalid questionnaires, such as those showing large sections of invariant responses, extremely short response times, or obviously illogical response patterns, will be excluded according to pre-specified quality control criteria.

For missing data, the proportion of missing values for each item and the overall level of missingness will first be reported. If the proportion of missing data is low, appropriate imputation strategies or available-case methods may be used in specific analyses. If an estimation method suitable for handling missing data is used in the confirmatory factor analysis stage, that method will be preferred. All missing-data handling procedures will be explicitly reported in the results.

Descriptive statistics for each item, including mean, standard deviation, skewness, and kurtosis, will also be examined to assess item dispersion and distributional characteristics and to identify potential floor or ceiling effects. If an item shows an overly concentrated distribution, insufficient variance, or weak discriminatory power, its retention will be evaluated in combination with subsequent item analysis results.

6.3 Stage 1: Item Analysis and Exploratory Factor Analysis

6.3.1 Item Analysis

In the first stage, item analysis will first be conducted on the 40 initial items to evaluate item discrimination and the association of each item with the overall scale. The specific procedures are as follows.

First, participants will be divided into the upper 27% and lower 27% groups based on the total scale score. Independent-samples t-tests will then be used to compare item scores between the high-score and low-score groups, and the critical ratio for each item will be calculated. In general, if an item has a critical ratio below 3, or if the difference between the high-score and low-score groups does not reach statistical significance ($P \geq 0.05$), the item will be considered to have insufficient discrimination and may be considered for deletion.

Second, Pearson correlation coefficients between each item and the total scale score will be

calculated. Items with item-total correlations below 0.35 will be considered weakly associated with the overall construct and may be considered for removal. For items with very low or directionally abnormal correlations with the total score, further judgment will be made in light of item content and theoretical relevance.

In addition, item communalities will be examined. If an item has a communality below 0.20, this will suggest that the item is not well explained by the latent factor structure and may be considered a candidate for deletion. Item deletion at this stage will not be based on any single statistical criterion alone; rather, it will be determined by integrating item discrimination, item-total correlation, communality, content representativeness, and theoretical relevance.

6.3.2 Suitability Testing for Exploratory Factor Analysis

After item analysis is completed, exploratory factor analysis will be conducted on the retained items. Before factor extraction, the suitability of the sample for factor analysis will be assessed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. In general, a KMO value greater than 0.70 and a statistically significant Bartlett's test ($P < 0.001$) will be considered evidence that the data are suitable for factor analysis.

6.3.3 Factor Extraction Method and Rotation

Because the purpose of this study is to identify latent construct structure rather than simply reduce data, exploratory factor analysis will be conducted using principal axis factoring or maximum likelihood extraction. As behavioral avoidance, cognitive avoidance, safety behavior, and emotion-driven behavior are theoretically expected to be correlated, an oblique rotation method, such as Promax or Oblimin, will be used to better reflect the underlying psychological structure.

If the items are rated using Likert-type response options and the data characteristics are more consistent with ordinal categorical variables, factor extraction methods based on an appropriate correlation matrix for ordinal data will be preferred. If the data approximate continuous distributions and show satisfactory distributional properties, this will be noted and justified in the interpretation of the results.

6.3.4 Determination of the Number of Factors

The number of factors will be determined based on multiple sources of evidence rather than relying solely on the eigenvalue-greater-than-one rule. Specifically, the decision will be based on parallel analysis, the scree plot, eigenvalues, the theoretical interpretability of the factors, and the distribution of items across factors. Priority will be given to a factor solution that is consistent with the theoretical framework and has clear psychological meaning; however, if the empirical results deviate from the theoretical expectations, decisions will be made cautiously by integrating statistical evidence and theoretical reasoning.

6.3.5 Item Retention Criteria

The following criteria will be used for item retention during the exploratory factor analysis stage: the primary loading of an item on its intended factor should be at least 0.45; if an item has relatively high loadings on two or more factors, the difference between the primary loading and the secondary loading should be greater than 0.20 to avoid substantial cross-loading; items with very low communalities, unclear factor membership, or obvious content redundancy will also be considered for deletion. In addition to statistical criteria, the completeness of content coverage within each dimension will also be taken into account to avoid compromising construct representativeness in pursuit of statistical optimization alone.

A stepwise item reduction strategy will be used. After each deletion, exploratory factor analysis will be rerun until a candidate version of the scale with a clear structure, well-defined item-factor relationships, satisfactory statistical indicators, and sound theoretical interpretability is obtained. At the end of the first stage, a preliminary structural model and a candidate set of items for the Excessive Emotional Behavior Scale will be established.

6.4 Stage 2: Confirmatory Factor Analysis

6.4.1 Independent-Sample Validation Strategy

After completion of the first stage, a new independent sample will be recruited for confirmatory factor analysis. This sample will not be involved in item analysis or exploratory factor analysis, thereby ensuring that the exploration and validation processes remain independent. The primary purpose of the second stage is to test the stability of the scale structure identified in the first stage and to evaluate the fit of alternative structural models.

6.4.2 Model Specification

Confirmatory factor analysis will first be used to test the four-factor structure identified in the exploratory factor analysis. If theoretically warranted, several competing models may also be compared, such as a one-factor model, a correlated four-factor model, and a higher-order factor model, in order to evaluate the relative fit of the target model. All model comparisons will be guided by theoretical considerations rather than purely data-driven model building.

6.4.3 Estimation Method

If the items use Likert-type response options and are more appropriately treated as ordinal categorical variables, an estimation method suitable for ordinal data, such as WLSMV, will be preferred. If the item distributions approximate continuous variables and meet the relevant statistical assumptions, maximum likelihood estimation may also be used. The final estimation method will be determined based on the response format, distributional characteristics, and model convergence, and this choice will be clearly reported in the results.

6.4.4 Model Fit Evaluation Criteria

Model fit will be evaluated using multiple fit indices, including the chi-square to degrees of freedom ratio (χ^2/df), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). In general, CFI and TLI values greater than 0.90 will indicate acceptable model fit, and values greater than 0.95 will indicate good fit; RMSEA and SRMR values below 0.08 will indicate acceptable fit, and values below 0.06 will indicate good fit. Fit indices will be interpreted comprehensively rather than mechanically based on a single cutoff.

6.4.5 Principles for Model Modification

If the initial model does not achieve satisfactory fit, limited model modifications may be considered based on modification indices. However, all modifications must be supported by clear theoretical justification, such as residual correlations arising from semantically similar items or items whose content deviates from the intended dimension. Overfitting based solely on modification indices will be avoided in order to preserve the theoretical interpretability and structural stability of the model.

6.5 Stage 3: Reliability and Validity Testing

6.5.1 Internal Consistency Reliability

After the formal structure of the scale has been established in the second stage, internal consistency will be evaluated for the total scale and each dimension. Reliability indices will include Cronbach's alpha and McDonald's omega. In general, alpha and omega values of 0.70 or higher will be considered acceptable, and values of 0.80 or higher will indicate good internal consistency.

During final scale formation, changes in internal consistency after deleting individual items, item-total correlations, and factor loadings will also be examined. However, item retention will not be determined solely on the basis of any single statistical index; instead, decisions will be made by integrating theoretical necessity, content representativeness, and the overall stability of the scale structure.

6.5.2 Convergent Validity

Convergent validity will be assessed using composite reliability (CR) and average variance extracted (AVE). In general, CR values of 0.70 or above and AVE values of 0.50 or above will be considered indicative of satisfactory convergent validity. If the AVE of an individual dimension is slightly below the recommended threshold but CR is adequate and the item content has clear theoretical value, the results will be interpreted in light of the overall structure and relevant literature.

6.5.3 Discriminant Validity

Discriminant validity will be evaluated by comparing the square root of the AVE for each

dimension with the correlations among dimensions. If feasible, the heterotrait-monotrait ratio (HTMT) will also be used. If the square root of the AVE for a given dimension exceeds its correlations with other dimensions, or if the HTMT is below commonly recommended thresholds, this will be considered evidence of adequate discriminant validity.

6.5.4 Criterion-Related Validity

If external variables theoretically related to excessive emotional behavior are measured in the study, criterion-related validity will be further examined through correlation analyses. For example, correlations may be examined between the Excessive Emotional Behavior Scale and constructs such as anxiety, depression, perceived stress, experiential avoidance, difficulties in emotion regulation, mindfulness, or resilience. Theoretically, the total score and relevant dimension scores of the Excessive Emotional Behavior Scale are expected to be positively correlated with anxiety, depression, stress, and experiential avoidance, and may be negatively correlated with mindfulness, cognitive flexibility, or resilience. Findings consistent with these expectations will provide further support for the validity of the scale.

6.6 Principles for Final Item Selection

The final number of items in the formal scale will not be fixed in advance but will be determined after completion of item analysis, exploratory factor analysis, confirmatory factor analysis, and reliability and validity testing. Final item retention will follow the following principles: first, items should show satisfactory statistical performance, including adequate factor loadings, reasonable communalities, and a positive contribution to internal consistency; second, item content should accurately reflect the core meaning of the intended dimension; third, each dimension should retain a sufficient number of items to ensure structural stability and content representativeness. In principle, each dimension should retain no fewer than three items to ensure the basic stability of latent variable measurement.

6.7 Statistical Software and Significance Level

Data analyses may be conducted using SPSS, R, Mplus, or Amos. Item analysis and initial exploratory factor analysis may be performed in SPSS or R, whereas confirmatory factor analysis, composite reliability, average variance extracted, and competing model comparisons are recommended to be conducted in Mplus, R, or Amos. All statistical tests will be two-tailed, and the significance level will be set at $\alpha = 0.05$.

6.8 Principles for Reporting Results

The study results will be reported in sequence, including sample characteristics, item analysis results, suitability testing and factor extraction results for exploratory factor analysis, the item screening process, model fit indices from the independent-sample confirmatory factor analysis, competing model comparison results, and the reliability and validity indices of the total scale and each dimension. For each step involving item deletion, model modification, and structural

adjustment, both the statistical rationale and the theoretical justification will be reported to ensure transparency of the research process and rigor in the interpretation of the findings.

7. Ethical Considerations

All participants will provide written informed consent prior to participation in the study. The study will be conducted in accordance with institutional ethical guidelines and relevant ethical standards for research involving human participants. Participant confidentiality and data security will be maintained throughout the study process.

8. Innovations

This study has several key innovations. First, it will provide the first systematic operationalization of excessive emotional behaviors within a transdiagnostic framework. Second, it integrates the theoretical perspectives of the Unified Protocol and Mindfulness Intervention for Emotional Distress, thereby enhancing both theoretical coherence and clinical applicability. Third, the study adopts a mixed-methods approach to scale development, combining qualitative and quantitative methods in order to improve the ecological validity, comprehensibility, and psychometric robustness of the final instrument.