

Effectiveness of Mentalization-Based Treatment in Reducing Suicidal Ideation Among Middle-Aged Women: A Quasi-Experimental Study

Salma. Maleki¹, Narjes Sadat. Mirkhan^{2*}, Bahareh. Ajdarbin¹

¹ Department of Psychology, WT.C., Islamic Azad University, Tehran, Iran

² Department of Psychology, SR.C., Islamic Azad University, Tehran, Iran

* Corresponding author email address: narjesalsadatmirkhan@iau.ac.ir

E d i t o r	R e v i e w e r s
Donatella Di Corrado [✉] Associate Professor, Department of psychology, Kore University of Enna, Enna, Italy donatella.dicorrado@unikore.it	Reviewer 1: Ali Khodaei [✉] Department of Psychology, Faculty of Educational Sciences and Psychology, Payam Noor University, Tehran, Iran. Email: alikhodaei@pnu.ac.ir Reviewer 2: Kamdin Parsakia [✉] Department of Psychology and Counseling, KMAN Research Institute, Richmond Hill, Ontario, Canada. Email: kamdinparsakia@kmanresce.ca

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Study Design and Participants section, the manuscript states that “eligible participants were assigned to the intervention and control conditions using simple random assignment,” yet the study is repeatedly described as quasi-experimental. Simple random allocation after eligibility screening would ordinarily move the design closer to a randomized controlled trial, provided that allocation was genuinely random and adequately concealed. Please clarify exactly how random assignment was performed, who generated the allocation sequence, whether allocation concealment was used, and whether the person enrolling participants could predict group assignment. The study design should then be labeled consistently throughout the title, abstract, Methods, Results, and Discussion as either randomized, quasi-experimental, or nonrandomized according to the actual procedure.

The sentence “To compensate for potential attrition during the intervention and follow-up phases, 36 women were initially recruited” needs additional information about the six participants who did not enter the final analysis. Please report the number and reasons for attrition separately by group, indicate whether dropouts occurred before treatment, during treatment, or before follow-up, and state whether withdrawal was related to adverse events, symptom worsening, scheduling problems, or loss to follow-up. Because suicidal ideation is the primary outcome, attrition associated with symptom exacerbation could

systematically bias the observed treatment effect. A CONSORT-style participant-flow diagram would be highly desirable even if the study is ultimately classified as quasi-experimental.

The Results paragraph stating that the MBT group improved from 28.47 at pretest to 13.07 at posttest and 12.40 at follow-up raises a clinically important question about baseline severity and floor effects. Participants began with scores described by the manuscript as indicating “serious suicidal intent,” yet posttreatment scores remained within a range the manuscript describes as “preparation for suicide.” The Discussion should therefore distinguish statistical improvement from clinical remission or safety. A large reduction in BSS score does not necessarily mean that participants ceased to require suicide-risk monitoring. Please report, if available, the proportion of participants crossing clinically meaningful thresholds or achieving reliable change, while avoiding unsupported claims of recovery.

The Results should provide more complete descriptive and inferential reporting at each measurement occasion. Table 1 currently includes means and standard deviations only. Given the small sample, please consider reporting medians and ranges or confidence intervals as supplementary descriptive information, especially because parametric normality tests have low power in groups of 15 participants. The statement that “Shapiro–Wilk tests were nonsignificant” should not be treated as proof of normality. Visual inspection of Q–Q plots and evaluation of influential observations would be useful, particularly because a small number of extreme scores could substantially affect the very large effect estimates.

The manuscript reports both Shapiro–Wilk and Lilliefors-corrected Kolmogorov–Smirnov tests, along with skewness and kurtosis, but the assumption-testing section does not address independence, influential observations, or equality of variance for the calculated change scores used in the independent-samples *t* tests. Because change scores became the principal inferential outcome after the ANCOVA assumption violation, their assumptions should be assessed directly. Please report Levene’s tests for the two change variables and indicate whether any outliers were identified. If variance equality was not met, Welch’s *t* test would be preferable to the pooled-variance *t* test.

Authors revised the manuscript and uploaded the document.

1.2. Reviewer 2

Reviewer:

The statement that “the minimum sample size of 15 participants per group was determined with consideration of a statistical power of 0.80 and an alpha level of 0.05” is insufficient for reproducibility. A formal power calculation also requires specification of the expected effect size, statistical test, number of groups, number of repeated measurements or covariates, and whether the calculation was one- or two-tailed. Please report the software used for the calculation, the assumed effect size and its empirical or theoretical justification, and the resulting minimum sample before allowance for attrition. Given the exceptionally large observed effects, a transparent *a priori* power analysis is particularly important to distinguish planned sample adequacy from post hoc justification.

The safety procedures require much more detailed reporting because the study enrolled women with substantial suicidal ideation, with baseline means around 28 on a 0–38 scale. The Methods currently state only that participants would be excluded if they “experienced symptom exacerbation or an acute psychological crisis requiring immediate clinical intervention.” Please specify the suicide-risk management protocol, including how acute risk was assessed, who performed the assessment, the criteria for emergency referral, whether safety plans were created, whether emergency contacts or psychiatric services were available, and how risk was monitored during online sessions and between sessions. For a study involving suicidal participants, these procedures are not ancillary methodological details but central ethical and clinical requirements.

The Data Collection Tool paragraph reports that BSS scores of “0–5 indicated suicidal ideation, 6–19 indicated preparation for suicide, and 20–38 represented serious suicidal intent.” This categorical interpretation should be carefully verified against the original or validated Persian scoring source because the Beck Scale for Suicide Ideation is most commonly interpreted primarily as a continuous severity measure, and threshold conventions vary substantially across studies. Please provide the precise validation source supporting these cut points and explain why they were appropriate for this population. If the cutoffs are not firmly validated, the manuscript should avoid presenting them as universally established clinical categories.

The BSS psychometric paragraph should distinguish between properties reported in previous studies and reliability obtained in the present sample. The manuscript currently states that previous research found Cronbach's alpha coefficients of 0.87–0.97 and that an Iranian study reported $\alpha = .95$, but no internal-consistency coefficient is provided for the current participants. Please calculate and report Cronbach's alpha or preferably omega for the BSS at baseline in the present sample. Although a sample of 30 produces an imprecise reliability estimate, reporting it would still provide useful information regarding measurement consistency in the actual dataset being analyzed.

The intervention paragraph states that MBT “was adapted for group delivery among middle-aged women,” but the nature and validation of this adaptation are unclear. Please explain whether the protocol was derived from a published MBT manual, whether adaptations were made by the research team, which elements were modified, and whether the adapted protocol underwent expert review, pilot testing, or content validation. Because the intervention incorporates topics such as menopause-related transitions, loss, loneliness, and self-harm thoughts, readers need to know which components belong to standard MBT and which were study-specific adaptations. Treatment reproducibility would also improve if the manuscript reported specific techniques used in each session rather than only broad thematic objectives.

The sentence that MBT “was delivered by a therapist with specific training in this therapeutic approach” is too vague for evaluation of treatment quality. Please report the therapist's professional discipline, academic degree, years of clinical experience, formal MBT training, supervision arrangements, and whether the therapist was also involved in recruitment or outcome assessment. In addition, the manuscript should report how treatment fidelity was monitored—for example, through session recordings, adherence checklists, independent ratings, or supervision. Without fidelity assessment, it is difficult to determine whether observed effects can confidently be attributed to delivery of MBT rather than to nonspecific therapist factors.

The control condition is described as receiving “no psychological intervention during the study period” and remaining on a waiting list. This design permits estimation of improvement relative to no treatment but cannot distinguish the specific effects of MBT from attention, expectancy, therapeutic alliance, group contact, or repeated interaction with a clinician. This limitation should be emphasized more prominently in the Discussion, particularly because the observed effect is extraordinarily large. The authors should avoid language implying that the study establishes the specific efficacy of MBT over other therapies and instead characterize the results as evidence of superiority to a waiting-list condition.

The statistical strategy requires reconsideration. The manuscript reports that “the group $\times$ pretest interaction used to assess the homogeneity of regression slopes was statistically significant, $F = 42.594$, $p < .001$,” after which ANCOVA was retained “for completeness” and supplemented with change-score t tests. Once the homogeneity-of-regression-slopes assumption is materially violated, the conventional ANCOVA group effect is not straightforwardly interpretable because the adjusted group difference depends on the covariate value. Rather than emphasizing an invalid adjusted main effect, the authors should consider a model that explicitly includes the group $\times$ baseline interaction or a longitudinal mixed-effects model using all three measurement occasions. The current analytical sequence risks presenting an extremely large ANCOVA effect that is statistically impressive but methodologically inappropriate.

The change-score analysis also warrants additional diagnostic information. Table 3 reports $t(28) = -38.488$ and $t(28) = -35.849$, which are exceptionally large test statistics for a behavioral intervention with $n = 15$ per group. Please verify the raw data, change-score standard deviations, and t -test calculations carefully. The manuscript should also report 95% confidence intervals for the between-group mean differences and standardized effect sizes such as Hedges' g . Given that the mean pretest-to-posttest changes differ by approximately 14.87 points while within-group SDs of change are only 1.40 and 0.52, the resulting standardized effect would be unusually large and merits explicit verification and transparent reporting.

Authors revised the manuscript and uploaded the document.

2. Revised

Editor's decision: Accepted.

Editor in Chief's decision: Accepted.