Shiraz E-Medical Journal
Volume 27 · Issue 4 · e168654

OPEN PEER REVIEW

Agreement Between Clinical Judgment and Standardized Questionnaires in Temperament (Mizaj) Assessment in Persian Medicine

Fereshteh Pirooz, Roshanak Ghods, Mohsen Saberi Isfeedvajani

SUBMITTED	REVISED	ACCEPTED	DOI
3 Dec 2025	2 Mar 2026	27 Mar 2026	10.5812/semj-168654

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Referee 1, Associate Editor, Editor-in-Chief 2
2 **Revision 1** Referee 1, Associate Editor, Editor-in-Chief 4
Authors' reply to reviewers 5
Reviewer and editor comments 11
3 **Revision 2** Referee 1, Associate Editor, Editor-in-Chief 13
Authors' reply to reviewers 14
Reviewer and editor comments 19
4 **Revision 3** Referee 1, Associate Editor, Editor-in-Chief 21
Authors' reply to reviewers 22
Reviewer and editor comments 26

1 Original submission

3 comments · 13 Dec 2025 – 6 Jan 2026

Referee 1 Reviewer

MAJOR REVISION 13 Dec 2025

Comments:
1. Revise the title. In this manuscript, agreement was assessed.
2. Talking about the Final Integrated Mizaj (FIM) based on consensus doesn't make sense. Please talk about the agreement among the three methods.
3. Why was clinical judgment performed by one person instead of a group of experts? It is not reliable.
4. It is not clear how the numerical scores from 0 to 2 were assigned.
5. What about the reliability and validity of each tool?
6. What do you mean by responses 1, 2, and 3?
7. What do you mean by responses 1 to 5 in Table 2?
8. Remove all information about FIM.
9. The sample size calculation is not based on the main aim (agreement).
10. Out of 437, 80 were excluded due to unwillingness? But unwillingness is not an exclusion criterion.
11. Report n (%) in Table 3.
12. The Chi-Square test is not appropriate for the paired data.
13. Present the results and discussion in two separate sections.
14. Table 4 is not clear.

Associate Editor

MAJOR REVISION 1 Jan 2026

This manuscript addresses an important and clinically relevant issue in Persian Medicine, namely the comparison between clinical judgment and standardized questionnaires for Mizaj assessment. The study benefits from a relatively large sample size and introduces an innovative concept (Final Integrated Mizaj, FIM). However, several methodological, structural, and reporting issues should be addressed to strengthen the scientific rigor and clarity of the manuscript.

1. The creation of FIM as a reference standard is innovative but requires stronger methodological justification.
2. Since FIM is derived from the same tools being evaluated, there is a potential risk of incorporation bias.
3. The rationale for prioritizing the Salmannezhad questionnaire when no agreement exists among the three methods should be more explicitly justified.
4. The abstract is informative but overly dense and contains excessive numerical detail. Consider reducing the number of reported Kappa values and prevalence percentages. The conclusion of the abstract should focus more clearly on the clinical implication of integrating clinical judgment with questionnaires.
5. The introduction provides a comprehensive background but is relatively long. The theoretical explanation of elemental qualities and Mizaj categories could be shortened. The research gap should be articulated more clearly and earlier in the introduction.
6. The clinical Mizaj assessment scoring system (0–2 scale) is mentioned, but the criteria for categorizing participants into Hot/Cold/Wet/Dry based on these scores are not fully explained. Please clarify whether inter-rater or intra-rater reliability of the physician's assessment was evaluated or acknowledge this as a limitation.
7. Tables 1 and 2 (full questionnaires) are very lengthy and may interrupt the flow of the manuscript. Consider moving the full questionnaires to supplementary materials and providing only a summarized description in the main text.

8. Interpretation thresholds for Kappa coefficients (e.g., poor, moderate, substantial agreement) should be briefly stated.
9. Some results are repeated verbatim in the Discussion section. The Results section should be strictly descriptive, while interpretation and comparison with other studies should be reserved for the Discussion.
10. Table 3 is difficult to read due to formatting and alignment issues. Column headings, abbreviations, and percentages should be presented more clearly for readability.
11. Several minor grammatical and typographical errors are present (e.g., “Resuls,” “Medcial Practice,” missing spaces such as “tonote”). A final language revision by a native or professional academic editor is recommended.
12. Some keywords overlap in meaning (e.g., “Iranian Traditional Medicine” and “Persian Medicine”). Consider refining the keywords to improve indexing and reduce redundancy.

Editor-in-Chief MAJOR REVISION 6 Jan 2026

Dear Author,
We have reviewed your manuscript carefully. This is to inform you that this manuscript needs major revision.
Kind Regards
EIC

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editorial Teams of Shiraz E-Med Journal,

We are pleased to careful and complete reviewed guides and comments. We tried to correct and write your comments and guides, carefully. We write our responses separately in two parts. For easy review, we add line numbers in the margin alongside of each line of the new manuscript and yellow highlight new words. We address with line numbers.

Response to Reviewers' Comments:

Part 1, Reviewer 0:

1. Revise the title. In this manuscript, the agreement was assessed.

We thanks and revise the title:

Concordance Between Clinical Judgment and Standardized Questionnaires in Temperament (Mizaj) Assessment in Persian Medicine: A Comparative Cross-Sectional Study.

2. Talking about the Final Integrated Mizaj (FIM) based on consensus doesn't make sense. Please talk about the agreement among the three methods.

That's right. We must talk about the agreement among the three methods instead of Integration. So we correct and re-write this title in article completely. In new manuscript from 148 to 169 lines.

The Final Agreed Mizaj (FAM) was defined as the Mizaj classification based on complete (100%) agreement .....

3. Why clinical judgment performed by a person instead of a group of experts? It is not reliable.

That's right. We must explain our aims. We try to present a practical way in clinic nearby actual. So we add lines 89 to 92 .
Given that the primary aim of the study was ...

4. It is not clear how the numerical score from 0 to 2 were assigned.

Nice . We must talk and clear how the numerical score from 0 to 2 were assigned. So we correct and re-write this part of article. In new manuscript from lines 99 to107 and lines 133 to138 and lines 165 to169.

For each participant, the physician evaluated both the active qualities ....

After completion of the two questionnaire-based assessments, ...

For each participant, eight data columns ....

5. How about the reliability and validity of each tool?

Those are very important. We missed them. So correct and add these from 111 to 114 lines in new manuscript.
The MMQ includes eight items (Questions 1–8) assessing Hotness–coldness .....

6. What do you mean by response 1, 2 and 3?

Thanks your nice comment. We add our mean by response 1, 2 and 3, from 114 to 116 lines in new manuscript.

Each item has three response options scored as 1(Cold in Questions 1–8, Wet …..

7. What do you mean by response 1 to 5 in Table 2?

Thanks your well comment. We add our mean by response 1 to 5 in Table 2, from 127 to 129 lines in new manuscript.

Each item has five response options scored as 1&2(Cold in Questions 1–15, Wet ….

8. Remove all of information about FIM.

Thanks for your nice attention; we removed those titles in article completely. All of them are yellow highlighted.

9. The sample size calculation is not based on the main aim (agreement).

This is correct. Because of being this study a part of PHD thesis ,that was approved by the Ethics Committee of Iran University of Medical Sciences (approval number: IR.IUMS.REC.1402.063-2023-04030), The sample size calculation was based on that. Line of 341.

10. Out of 437, 80 were excluded due to unwillingness? But unwillingness is not an exclusion criteria.

We excuse, that's right. We must write Exclusion criteria completely. We correct that in new manuscript from line 87 to 88 .

Exclusion criteria included unwillingness ….

11. Report n (%) in Table 3.

We thank, its nice comment and we do it completely in table1 . Because of deleting Tables1&2 from new manuscript, Table 3 change to table1 . After line 194 in new manuscript.

.

12. The Chi-Square test is not appropriate for the paired data.

It is mistake to write. We delete it.

13. Talk about the results and discussion in two different parts.

It has been corrected. We separate results and discussion in two different parts. Line 217 in new manuscript.

14. Table 4 is not clear.

We thank, it's nice comment and we do it completely in table1 . Because of deleting Tables1&2 from new manuscript, Table 4 changes to table2. New manuscript line 210.

Part 2, Associate Editor's Comments:

Associate Editor 0: This manuscript addresses an important and clinically relevant issue in Persian Medicine, namely the comparison between clinical judgment and standardized questionnaires for Mizaj assessment. The study benefits from a relatively large sample size and introduces an innovative concept (Final Integrated Mizaj, FIM). However, several methodological, structural, and reporting issues should be addressed to strengthen the scientific rigor and clarity of the manuscript.

1. The creation of FIM as a reference standard is innovative but requires stronger methodological justification.

We happy you are agreeing with us. That's right. We try to complete and strong our methodological justification. We re-write line 139 to 164 in new manuscript.

As expected, the three assessment methods …….

2. Since FIM is derived from the same tools being evaluated, there is a potential risk of incorporation bias.

That's right. So we agree with you to change FIM to Final Agreed Mizaj (FAM), and choose this new term. We complain and use this title in new manuscript (from line 133 to 169).

After completion of the two questionnaire-based assessments, ……

3. The rationale for prioritizing the Salmannezhad questionnaire when no agreement exists among the three methods should be more explicitly justified.

Nice. We must talk about this selection. In study we know that the results obtained from these three Mizaj assessment methods are not only inconsistent but even completely contradictory, presenting a substantial methodological challenge. So we do a pilot study and analysis that results. They show that for identifying warmness–coldness and dryness – wetness Mizaj, agreement between CMA and SMQ was higher than another. So we prefer SMQ results when no agreement exists among the three methods. Based on the results obtained from the pilot study and that the number of cases of 100% non-conformity of 3 methods of measuring temperament in active and passive was less than 10 cases, the researchers decided to consider the results of the Salmannezhad questionnaire as the individual's temperament in all cases of disagreement between them.

Due to the increase in the number of tables in the article, I had to remove this table, which is presented for the information of the respected referees, that if they wish, due to the removal of the first 2 tables, it is not possible to edit the main text of the article.

Table 3. Frequency and Percentage of Agreement between Mizaj Assessment Methods

Level of Agreement	Frequency	(%)
All three consistent in both active and receptive qualities Researcher	71	20.3
All three consistent only in active quality	82	23.4
All three consistent only in receptive quality.	67	19.1

Both questionnaires consistent in both active and receptive qualities	7	2
One questionnaire consistent with researcher in both active and receptive qualities	71	20.3
All three inconsistent in both active and receptive qualities	2	0.6
Other cases	50	14.3
total	350	100

(in new manuscript from line 143 to 164).

After enrollment of 200 participants (more than 50% of the total sample), ….

4. The abstract is informative but overly dense and contains excessive numerical detail. Consider reducing the number of reported Kappa values and prevalence percentages.

The conclusion of the abstract should focus more clearly on the clinical implication of integrating clinical judgment with questionnaires.

We revise and re-write it. The Methods and Results of abstract , the conclusion of abstract change with this path. In new manuscript from line 11 to 22 for the abstract and from line 24 to 27 for the conclusion of the abstract.

After providing written consent,……

Given that the integration of Mizaj assessment into community……

5. The introduction provides a comprehensive background but is relatively long. The theoretical explanation of elemental qualities and Mizaj categories could be shortened.

We want to show importance of Mizaj assessment but your sense is right and the introduction is relatively long. So we effort to cut that and write shorter usage documents that you suggest. Lines 41 to 51 in new manuscript.

Numerous external factors, including residential climate, … .

The research gap should be articulated more clearly and earlier in the introduction.

Nice. We effort to articulate and clear the research gap better and earlier than past . Lines 58 to 62 and 71 to 80 in new manuscript.

In some instances, however, the results …..

The need for simple, reproducible, ….

6. The clinical Mizaj assessment scoring system (0–2 scale) is mentioned, but the criteria for categorizing participants into Hot/Cold/Wet/Dry based on these scores are not fully explained.

We explain the criteria for categorizing participants into Hot/Cold/Wet/Dry based on scoring system (0–2 scale) for The clinical Mizaj assessment mention as well as possible and write it in new manuscript line 99 to 107.

For each participant, the physician evaluated both the active qualities (Kayfiyyāt-e Fāʿelah: …….

Please clarify whether inter-rater or intra-rater reliability of the physician's assessment was evaluated or acknowledge this as a limitation.

Thank you very much for your attention to detail. The expert physician who conducted the Mizaj assessment at the field (interviews with volunteers); had completed academic training before starting work at the Faculty of Persian Medicine of the University of Iran. She had completed more than 50 cases of successful the Mizaj assessment and 100% agreement with the panels of at least 3 professors of this center.

However, considering that the reliability and validity of the clinical Mizaj assessment method has not been calculated and reported statistically and separately; we must accept it as one of the limiting factors of the research.

We add it in new manuscript line 312 to 318.

7. Tables 1 and 2 (full questionnaires) are very lengthy and may interrupt the flow of the manuscript. Consider moving the full questionnaires to supplementary materials and providing only a summarized description in the main text.

It's the best suggestion and we agree whit you completely. So it's excluded and transport them to supplementary files. You can read the summarized description in the main text from 111 to 132 lines.

The MMQ includes eight items (Questions 1–8) assessing…..

8. Interpretation thresholds for Kappa coefficients (e.g., poor, moderate, substantial agreement) should be briefly stated.

Thank you for a nice notification. It's neglected and we add it in lines 214 to 216 .

In accordance with conventional interpretation of Cohen's kappa values, …..

9. Some results are repeated verbatim in the Discussion section. The Results section should be strictly descriptive, while interpretation and comparison with other studies should be reserved for the Discussion.

It's the best that re-write the Discussion section and that is written again with comment clues from line 217 to 304 in new manuscript.

5.Discussion

The present study aimed to identify a valid, simple…….

10. Table 3 is difficult to read due to formatting and alignment issues. Column headings, abbreviations, and percentages should be presented more clearly for readability.

We thank, it's nice comment and we do it completely in table1 .Because of deleting Tables1&2 from new manuscript, table 3 change to table1 . Line 194 of new manuscript.

11. Several minor grammatical and typographical errors are present (e.g., “Resuls,” “Medcial Practice,” missing spaces such as “tonote”). A final language revision by a native or professional academic editor is recommended.

We correct them in new manuscript and they are highlighted.

Medical Practice, Results, Footnotes.

12. Some keywords overlap in meaning (e.g., “Iranian Traditional Medicine” and “Persian Medicine”). Consider refining the keywords to improve indexing and reduce redundancy.

We want to consider some keywords (e.g., “Persian Medicine”) more than past . So we use it in abstract and delete from Key words.

Keywords: Persian Medicine; Integrative Medicine; Mizaj; Medcial Practice; Temperament.

2 Revision 1

3 comments · 25 Feb 2026 – 6 May 2026

Referee 1 Reviewer

MINOR REVISION 25 Feb 2026

The sample size calculation is not based on the main aim (agreement).

Associate Editor

MAJOR REVISION 6 May 2026

This manuscript addresses an important issue in Persian Medicine: the lack of agreement between Mizaj assessment tools. The topic is relevant to integrative medicine, and the study attempts methodological innovation by proposing an integrated agreement-based classification (FAM). However, the study has major methodological and conceptual limitations that currently prevent it from meeting the standards of a robust scientific publication.

The study assumes that Mizaj is a measurable and classifiable biological construct. However, no objective biological gold standard exists for Mizaj. The study compares subjective tools with each other, which creates a circular validation problem. FAM is derived from the same tools being evaluated. Then, agreement is measured against FAM! This is not validation but internal consistency analysis. Claims about "true Mizaj" or "accuracy" are scientifically unsupported. I recommend reframing the study as: "Agreement analysis between assessment methods" NOT: "Determination of most accurate method".

The authors define FAM using CMA + MMQ + SMQ. They then evaluate the agreement of these same methods with FAM. This introduces incorporation bias, which artificially inflates agreement (especially for CMA). You should not use FAM as a reference standard. Instead, you should report pairwise agreement only or define an external, independent reference (if possible).

CMA was performed by only one physician. There is no inter-rater reliability, intra-rater reliability, or blinding. The authors acknowledge this as a limitation, but it severely weakens conclusions that "CMA is closer to "true Mizaj"". You need to remove comparative superiority claims such as "accuracy" and "true Mizaj" and clearly describe CMA as a "single-observer subjective assessment".

There is a substantial gender imbalance in the population of healthcare workers in Tehran (75.7%

Public
Comments*

male), which implies that your sample is not representative of the general population. Also, there is an additional concern. "Unwillingness" is treated as an exclusion criterion in this study, which introduces self-selection bias. You should emphasize the limited external validity and avoid claims about the general population.

While you used Kappa, which is appropriate for assessing agreement, no confidence intervals or adjustments for prevalence bias are reported. In interpreting Kappa, you used standard thresholds but ignored very low values (e.g., 0.054) indicating near-random agreement. Also, analyses such as sensitivity/specificity (despite claims of diagnostic relevance), multivariate analysis, and agreement by subgroup

(sex, age) are missing. I recommend that you add confidence intervals, interpret low Kappa values more critically, and remove “diagnostic accuracy” language. Because of FAM and its major methodological flaws, I suggest that you either clearly predefine the algorithm (with justification) or remove FAM entirely.
The manuscript concludes that integration provides the “most accurate” Mizaj, but there is no gold standard, outcome validation, or biological correlation. Thus, this conclusion is not supported. It would be much better to conclude that “Combining methods increases internal agreement”.
Language issues persist in grammar and phrasing. Also, there is some redundancy between the Results and Discussion. The tables are still difficult to interpret. The abstract needs a clearer aim, main findings, and limitations.

Editor-in-Chief MAJOR REVISION 6 May 2026

Dear Author,
We have reviewed your manuscript carefully. This is to inform you that this manuscript needs major revision.
Kind Regards
EIC

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor,

We sincerely thank the reviewer for this insightful comment. We agree that the previous version may have implied more than what the agreement data could support. In the revised manuscript, we have removed all references to FAM and limited the analysis to the three pairwise comparisons among CMA, MMQ, and SMQ. We also report the 95% confidence intervals for all Kappa coefficients in Table 2 and interpret the results more cautiously in the Discussion, emphasizing that the observed agreement ranged from slight to moderate rather than suggesting diagnostic superiority of any method. In addition, we revised the Conclusion and Limitations sections to avoid any claims of accuracy or gold-standard status and to acknowledge the absence of an external reference standard. We believe these changes have made the manuscript more methodologically sound and balanced.

.... Reviewers' Comments:

1. The sample size calculation is not based on the main aim (agreement).

Response:
We agree with this observation. The manuscript has been revised to clarify that the sample size in the parent study was calculated to estimate associations rather than to assess agreement statistics. Accordingly, the agreement analysis reported in the present study represents a secondary analysis of the same dataset. This point has been explicitly stated in both the Methods and the Limitations sections, and the reported Kappa values have been interpreted with this consideration.

... Associate Editor's Comments:

1. FAM is derived from the same tools being evaluated. Then, agreement is measured against FAM! This is not validation, but internal consistency analysis. Claims about “true Mizaj” or “accuracy” are scientifically unsupported. I recommend to reframe the study as:“Agreement analysis between assessment methods” NOT: “Determination of most accurate method”.

Response:

The title has been changed to: “*Agreement Between Clinical Judgment and Standardized Questionnaires in Temperament (Mizaj) Assessment in Persian Medicine*”. In addition, all claims regarding “true Mizaj” or “accuracy” have been removed throughout the manuscript.

2. The authors define FAM using CMA + MMQ + SMQ. Then, evaluating agreement of these same methods with FAM. This introduces incorporation bias which causes artificial inflation of agreement (especially for CMA). You should not use FAM as a reference standard. Instead,

you should report pairwise agreement only or define an external independent reference (if possible).

Response:

Thank you for this important comment. Following the reviewer's suggestion, we revised the analysis approach and removed the Final Agreed Mizaj (FAM) variable from the manuscript. Since the study focuses on agreement between assessment methods, the analyses were redesigned to report only pairwise comparisons between the three original methods (CMA, MMQ, and SMQ). This approach avoids the potential for incorporation bias and allows the agreement between methods to be evaluated directly.

3. CMA performed by one physician only. There is no inter-rater reliability, intra-rater reliability, or blinding. Authors acknowledge this as limitation, but this severely weakens conclusions that "CMA is closer to "true Mizaj"". You need to remove comparative superiority claims such as "accuracy" and "true Mizaj" and clearly state CMA as a"single-observer subjective assessment".

Response:

As stated above, all claims related to "true Mizaj" or "accuracy" have been removed from the manuscript, and no comparative superiority of CMA is asserted. It is also important to clarify that the physician's Mizaj assessment was performed prior to completion of the self-report questionnaires by the participants. Consequently, the physician was unaware of the questionnaire-based Mizaj classifications at the time of clinical assessment. While this does not address inter-rater reliability, it does represent a form of assessor blinding that reduces potential information bias and increases confidence in the independence of the CMA findings.

4. There is great gender imbalance in the population of healthcare workers in Tehran (75.7% male) which implies that your sample is not representative of general population.

Response:

We appreciate the reviewer's comment regarding the gender imbalance within our sample. We acknowledge that the predominantly male composition of the healthcare workforce in Tehran limits the representativeness of our study population. This issue has been taken into careful consideration throughout the manuscript, and we have interpreted the findings conservatively in light of this limitation. We have also explicitly noted this point in the Limitations section.

5. "Unwillingness" treated as exclusion criterion in this study which introduces self-selection bias. You ought to emphasize limited external validity and avoid general population claims, surely.

Response:

Thank you for this valuable comment. We would like to clarify that our intention in the original manuscript was not to define unwillingness to continue participation as a methodological exclusion criterion, but simply to highlight that participation in the study was entirely voluntary. To prevent any misunderstanding, we have revised the Methods section and limited the exclusion criteria solely to scientific factors that could interfere with the assessment process.

6. While you Used Kappa that is appropriate for agreement but, there are no confidence intervals and no adjustment for prevalence bias reported. During the interpretation of Kappa, you used standard thresholds but ignored very low values (e.g., 0.054) indicating near-random agreement. Also, there are missing analyses such as Sensitivity/specificity (despite claiming diagnostic relevance), Multivariate analysis, Agreement by subgroups (sex, age). I recommend you to add confidence intervals, interpret low Kappa more critically, and remove “diagnostic accuracy” language. Because of FAM and its majormethodological flaws, I suggest you either predefining algorithm clearly (with justification) or remove FAM entirely.

Response:

We appreciate the reviewer’s valuable suggestion. In response, we revised the manuscript substantially. All references to FAM were removed, and the analysis now focuses exclusively on the pairwise agreement between CMA, MMQ, and SMQ. We ensured that the 95% confidence intervals are reported for all Kappa estimates in Table 2, and we revised the Discussion to interpret low Kappa values more critically and conservatively. We also removed any wording that could be interpreted as implying diagnostic accuracy or superiority of one method over another. Finally, we expanded the limitations to highlight the absence of an independent gold standard and the context-specific nature of the agreement findings. These revisions, we hope, address the reviewer’s concern appropriately.

7. The manuscript concludes: integration provides “most accurate” Mizaj but, there are no gold standard, no outcome validation, and no biological correlation. So, this conclusion is not supported. It is so better to conclude that “Combining methods increases internal agreement”.

Response:

Thank you for this important comment. We agree that, without a gold standard, outcome validation, or biological correlates, it is not appropriate to claim that any method provides the “most accurate” assessment of Mizaj.

Accordingly, the Discussion and Conclusion were revised to emphasize that this study examines the level of agreement between commonly used Mizaj assessment methods rather than their diagnostic accuracy. The conclusion now focuses on the observed agreement between the

questionnaires and clinical assessment, and indicates that combining methods may improve internal consistency rather than define a definitive or most accurate standard.

8. Language issues persist in grammar and phrasing.

Response:

English polish has been done.

9. Some redundancy are between Results and Discussion.

Response:

The whole article has been edited with considering all these issues.

10. Tables are still difficult to interpret.

Response:

Tables were revised.

11. Abstract needs clearer aim,main findings, and limitations.

Response:

Abstract has been fully revised according to the journal standards.

and

All 3 suggested papers were added to the article (Ref No.15.19,and 22).

I apologize for any shortcomings that remained in the previous version and would be grateful for any further comments that may help us refine the manuscript.

Sincerely yours,

Prof. Roshanak Ghods
Corresponding author

3 Revision 2

3 comments · 13 May 2026 – 31 May 2026

Referee 1 Reviewer

MINOR REVISION 13 May 2026

The sample size calculation is not clear. Please justify the sample size based on information from a previous study and provide a reference.

Associate Editor

MAJOR REVISION 31 May 2026

The authors did make meaningful revisions. These revisions improve readability and reduce some interpretive overreach, but the core methodological problem is not solved. FAM is still derived from the same tools being evaluated. Even after renaming FIM → FAM, the problem persists.
Agreement with FAM cannot be interpreted as evidence of validity, superiority, or diagnostic correctness. At best, the study shows internal concordance patterns among three non-independent methods. The manuscript still does not clearly acknowledge this limitation.
The justification for prioritizing the Salmannezhad questionnaire in disagreement cases is not convincing. You state that you performed a pilot study and selected the questionnaire with higher agreement with CMA. This creates multiple concerns, including potential data-driven post hoc decision-making, risk of confirmation bias, no predefined algorithm, and no external validation. Additionally, the pilot data are not presented formally, no statistical justification is shown, and the decision appears arbitrary. This significantly weakens reproducibility.
The authors justify using one physician because they wanted a "practical clinic-like approach." However, this does not address the scientific concern about the lack of inter-rater or intra-rater reliability. For a study fundamentally about diagnostic agreement, this is a substantial limitation.
The authors admit "the sample size was based on the PhD thesis approval rather than agreement analysis". This is not an acceptable methodological justification in a scientific manuscript. Agreement studies generally require kappa-based sample size calculations, precision estimates, or expected concordance assumptions. This item remains unresolved.
Despite the concerns, the study still has value. The manuscript should clearly state that this is an exploratory agreement/concordance study only. This is NOT a validation, superiority, integration, or "true Mizaj" identification study.
The authors must explicitly discuss incorporation bias. This limitation must appear clearly in Methods, Discussion, and Limitations.
The authors should remove preferential interpretation of FAM. FAM should be presented only as an operational exploratory construct, NOT as a reference standard.
The manuscript should explain the pilot analysis formally, justify the decision algorithm, and acknowledge post hoc decisions. You must add a stronger limitations section that includes single physician assessment, no inter-rater reliability, no gold standard, sample size limitations, and selection bias.

Editor-in-Chief

MAJOR REVISION 31 May 2026

Dear Author,
We have reviewed your manuscript carefully. This is to inform you that this manuscript needs major revision.

Kind Regards
EIC

4

Revision 3

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Responses to Reviewer Comments

Reviewer (1) Comments:

The sample size calculation is not clear. Please justify the sample size based th information of a previous study with reference.

We thank the reviewer for raising this important methodological point. Although the sample size was determined by the parent study, a post-hoc adequacy check using the formula proposed by Donner & Eliasziw (1992) (Reference 18 was added) demonstrated that N = 350 is sufficient to estimate a Kappa of 0.6 with a precision of ±0.15 at the 95% confidence level (required n = 349). This precision is sufficient to distinguish moderate from fair agreement, which is the clinically relevant threshold in this context. This has been added to the Methods section.

Associate Editor's Comments:

1. The authors did make meaningful revisions. These revisions improve readability and reduce some interpretative overreach but, the core methodological problem is not solved. FAM is still derived from the same tools being evaluated even after renaming FIM → FAM, the problem persists.

Response:

We respectfully note that the concept of "FAM" (Final Agreement Method) was entirely removed from the manuscript in the second revision. The revised manuscript presents only pairwise comparisons between the two questionnaires (SMQ and MMQ) and each questionnaire against the physician's clinical assessment (CMA), with no reference to FAM anywhere in the text. It appears that this comment may have been based on a previous version of the manuscript rather than the current revised submission. We kindly request the reviewer to re-evaluate this concern in light of the second revised manuscript.

2. Agreement with FAM cannot be interpreted as evidence of validity, superiority, or diagnostic correctness. At best, the study shows internal concordance patterns among three non-independent methods. The manuscript still does not clearly acknowledge this limitation.

Response:

As detailed in our response to Comment 1, the concept of FAM was entirely removed from the second revised manuscript. Consequently, this concern does not apply to the current version. The manuscript presents pairwise agreement analyses only, and we have explicitly acknowledged that these findings reflect concordance patterns among methods rather than diagnostic validity or superiority. We respectfully refer the reviewer to the revised manuscript for confirmation.

3. The justification for prioritizing the Salmannezhad questionnaire in disagreement cases is not convincing. You state that you performed a pilot study and selected the questionnaire with higher agreement with CMA. This creates multiple concerns such as: potential data-driven post hoc decision-making, risk of confirmation bias, no predefined algorithm, and no external validation.

Additionally; the pilot data are not presented formally, no statistical justification is shown, and the decision appears arbitrary. This significantly weakens reproducibility.

Response:

We thank the reviewer for this detailed critique. However, as noted in our responses to Comments 1 and 2, **the concept of FAM and the associated decision-making framework, including any prioritization of one questionnaire over another in disagreement cases, were entirely removed from the second revised manuscript.**

The current version does not employ any tiebreaking algorithm, does not prioritize the SMQ over the MMQ in disagreement cases, and makes no reference to a pilot study. The revised manuscript presents straightforward pairwise agreement analyses between SMQ, MMQ, and the physician's clinical assessment (CMA), without any hierarchical weighting of methods.

Consequently, the concerns raised in this comment, including post hoc decision-making, confirmation bias, absence of a predefined algorithm, and lack of external validation of a prioritization rule, do not apply to the current submission.

We respectfully refer the reviewer to the second revised manuscript for confirmation that these elements are no longer present in the text.

4. The authors justify using one physician because they wanted a "practical clinic-like approach." However, this does not address the scientific concern of lack of inter-rater or intra-rater reliability. For a study fundamentally about diagnostic agreement, this is a substantial limitation.

Response:

We thank the reviewer for this important methodological point. We acknowledge that CMA was performed by a single PM specialist, which precludes estimation of inter-rater reliability. We have revised the Limitations section to explicitly state that "*future multi-rater studies should address this gap*."

Furthermore, we clarify that because no independent external gold standard for mizaj assessment currently exists, CMA serves as a reference comparator rather than a validated criterion standard. This distinction is now explicitly stated in the revised Limitations: "*findings should be interpreted as reflecting agreement between methods rather than diagnostic validity*."

5. The authors admit "the sample size was based on the PhD thesis approval rather than agreement analysis". This is not an acceptable methodological justification in a scientific manuscript. Agreement studies generally require kappa-based sample size calculations, precision estimates, or expected concordance assumptions. This item remains unresolved.

Response:

We thank the reviewer for raising this important methodological point. Although the sample size was determined by the parent study, a post-hoc adequacy check using the formula proposed by Donner & Eliasziw (1992) demonstrated that N = 350 is sufficient to estimate a Kappa of 0.6 with a precision of ±0.15

at the 95% confidence level (required n = 349). This precision is sufficient to distinguish moderate from fair agreement, which is the clinically relevant threshold in this context. This has been added to the Methods section.

6. Despite the concerns, the study still has values. The manuscript should clearly state that this is an exploratory agreement/concordance study only. This is NOT a validation, superiority, integration, or "true Mizaj" identification study.

Response:

We agree with the reviewer's concern. The revised Limitations section now explicitly states: "*Because no independent external gold standard for mizaj assessment exists, findings should be interpreted as reflecting agreement between methods rather than diagnostic validity*." This framing is also consistent with the study's stated objective as an exploratory inter-method agreement analysis, as clarified in the Introduction (Objective section).

7. The authors must explicitly discuss incorporation bias. This limitation must appear clearly in Methods, Discussion, and Limitations.

Response:

Although CMA was performed by the same specialist whose clinical framework informed the study design, the assessments were conducted prior to and independently of questionnaire completion, ensuring that clinical judgments were not influenced by questionnaire responses. Therefore, while no external gold standard exists, the comparator retains methodological independence within the study design.

8. The authors should remove preferential interpretation of FAM. FAM should be presented only as an operational exploratory construct, NOT as a reference standard.

Response:

As addressed in our responses to Comments 1–3, **the FAM concept was entirely removed in the second revision.** The current manuscript compares only three methods pairwise (CMA, MMQ, SMQ), with no method designated as a reference standard.

9. The manuscript should explain the pilot analysis formally, justify the decision algorithm, and acknowledge post hoc decisions. You must add stronger limitations section including: single physician assessment, no inter-rater reliability, no gold standard, sample size limitations, and selection bias.

Response:

We respectfully note **that our manuscript does not claim to have conducted a pilot study at any point. No pilot phase was described or implied in any version of the manuscript.** Therefore, we

are unable to provide a description of a pilot study that was not performed. If the reviewer interpreted any part of the text as suggesting a pilot phase, we would welcome clarification of the specific passage, so we may address any ambiguity.

Prof. Roshanak Ghods
Corresponding Author

4 Revision 3

3 comments · 4 Jun 2026 – 7 Jun 2026

Referee 1 Reviewer

ACCEPT 4 Jun 2026

Accept

Associate Editor

ACCEPT 7 Jun 2026

The manuscript is suitable for publication in its current form. Congratulations on your work.

Editor-in-Chief

ACCEPT 7 Jun 2026

Dear Authors,
This is to inform you that your manuscript has been accepted for publication.
Kind Regards
EIC