Flores-Cohaila, J. A., Miranda-Chavez, B., Huarcaya-Victoria, J., Copaja-Corzo, C., Montesinos, S. Y., & Estela-Fernandez, C. A. (2026). Evidence of validity of the 10-item Big Five Inventory (BFI-10) in Peruvian medical students. Interacciones, 12, e542. https://doi.org/10.24016/2026.v12.542

LETTER OF REVIEWERS

Reviewer D:
Recommendation: Revisions Required

Relevance: High

Novelty: Moderated

Presentation and writing: Moderated

Comments for authors: Be as accurate as possible when making your comments. List each recommendation so that it is easy for authors to respond appropriately to each one. Indicate in a timely manner where changes should be made (i.e. paragraph 2 of the method section).
This study reports a psychometric validation of the BFI-10 in 248 Peruvian medical students, including CFA, measurement invariance by gender and academic phase, internal consistency, and limited criterion validity. The topic is relevant and the general structure is adequate. However, several critical issues compromise the soundness of the conclusions, particularly a factor-labeling error in the main CFA table, implausible inter-factor correlations, and a criterion validity section that is insufficient to support the validity claims made in the abstract and conclusions.

MAJOR COMMENTS:

Comment 1. Critical factor mislabeling in Table 3
The item content in Table 3 does not match the factor labels. Items BF5–BF6 (creativity, inventiveness) are assigned to Conscientiousness; BF7–BF8 (efficiency, hard-working) to Neuroticism; and BF9–BF10 (tension, nervousness) to Openness — all three assignments are inverted relative to their content. This is consistent with the polychoric matrix in Table 2, where BF9–BF10 show near-zero correlations with all other items and a high inter-item correlation (r = 0.63), the pattern expected of Neuroticism items, not Openness. The authors must correct the table and verify that all interpretations, external validity analyses, and invariance conclusions refer to the correctly labeled factors.

Comment 2. Implausible inter-factor correlations
The inter-factor correlations reported in Table 3 include values of 0.84 (Conscientiousness–Openness) and 0.75 (Agreeableness–Conscientiousness), which are inconsistent with the Big Five literature and suggest potential model misspecification. A likely contributor is the equality constraint imposed on factor loadings within each factor — a necessary identification strategy for two-indicator factors but one that can distort factor covariances. The authors should acknowledge this explicitly and discuss whether these correlations can be interpreted as substantively meaningful or are an artifact of the constrained model.

Comment 3. Borderline configural fit for gender undermines invariance conclusions
The configural model for gender (Table 4) yields CFI = 0.930 and RMSEA = 0.092. The RMSEA exceeds the authors' own acceptability threshold (<0.08), and the CFI barely clears the minimum. Scalar invariance derived from a poorly-fitting baseline model cannot be considered robust. The conclusions regarding gender invariance should be substantially qualified.

Comment 4. Criterion validity evidence is overstated
The only significant external validity finding is r = 0.13 between conscientiousness and GPA (< 2% shared variance), with no significant result for the neuroticism–psychiatric history comparison. These results constitute very weak criterion validity evidence. The abstract and conclusions should be revised to accurately reflect this: the BFI-10 shows adequate structural validity, but criterion validity evidence is insufficient and preliminary.

MINOR COMMENTS:

Comment 5. In the Variables section, "the ultra-short version reported by Costa [ref. 8]" should reference Rammstedt & John (2007) [ref. 6], which is the original BFI-10 paper. Reference 8 is a Brazilian validation study.

Comment 6. The term "responsibility" used in the Results and Discussion should be "conscientiousness" throughout.

Comment 7. Table 4 headers and labels are in Spanish; they should be translated to English for consistency with the rest of the manuscript.

Reviewer E:
Recommendation: Revisions Required

Relevance: Moderated

Novelty: Moderated

Presentation and writing: Moderated

Comments for authors: Be as accurate as possible when making your comments. List each recommendation so that it is easy for authors to respond appropriately to each one. Indicate in a timely manner where changes should be made (i.e. paragraph 2 of the method section).
On page 3, in the second paragraph, it would be important to include examples of evidence derived from personality assessments conducted in Latin America, as well as examples of brief instruments that have failed to demonstrate robust psychometric properties, in order to further justify the need for the present study. Additionally, the discussion of the available validity and reliability evidence for the BFI-10 should be expanded, as this would facilitate a more comprehensive comparison with the findings obtained in the current sample.
It is also necessary to support, with more recent references, the rationale for considering two items sufficient to assess personality traits, particularly given that several studies recommend a minimum number of items for adequate construct measurement.
In the title of Table 1, the term “sample” should be used instead of “population,” as the table refers to sample characteristics rather than population parameters. Additionally, the tables are not formatted according to APA guidelines and should be revised accordingly. Regarding the inter-item relationships, it is recommended to clearly identify which items correspond to each dimension in order to verify that the correlations between the two items within each dimension are sufficiently high, while correlations with items from other dimensions remain low. This criterion is particularly important for supporting the adequacy of representing each factor with only two items.
In Table 3, Items 7–10 appear to be misclassified, as the items assigned to openness actually correspond to neuroticism, and vice versa.
Measurement invariance was evaluated across sex, not gender; consequently, the terminology used in Table 4 should be modified to reflect this distinction accurately.
The categories presented in Table 4 are written in Spanish rather than English and should therefore be translated, along with the corresponding table footnotes.
The dichotomous indicator of prior psychological or psychiatric treatment may be insufficient for establishing external validity evidence. More detailed measures addressing the timing of treatment and associated symptomatology are necessary to adequately support this type of validity evidence. Accordingly, it is recommended that convergent and divergent validity evidence be examined using additional variables.
The categorization of academic average scores into intervals before examining their associations with personality dimensions is not recommended when exact continuous scores are available, as this procedure may result in a loss of information.
The manuscript is currently formatted using a Vancouver-style referencing system rather than the APA format required by the journal guidelines. Both the in-text citations and the reference list should therefore be revised to ensure consistency with APA style throughout the document.

Flores-Cohaila, J. A., Miranda-Chavez, B., Huarcaya-Victoria, J., Copaja-Corzo, C., Montesinos, S. Y., & Estela-Fernandez, C. A. (2026). Evidence of validity of the 10-item Big Five Inventory (BFI-10) in Peruvian medical students. Interacciones, 12, e542. https://doi.org/10.24016/2026.v12.542

Reviewer F:
Recommendation: Revisions Required

Relevance: High

Novelty: Moderated

Presentation and writing: High

Comments for authors: Be as accurate as possible when making your comments. List each recommendation so that it is easy for authors to respond appropriately to each one. Indicate in a timely manner where changes should be made (i.e. paragraph 2 of the method section).
It is important to acknowledge the authors' effort to address the validation of a personality assessment instrument. The manuscript is noteworthy for its rigorous use of factor analysis and measurement invariance testing, which provide relevant evidence regarding internal structure and fairness across groups. However, the methodological rigor of the manuscript should be strengthened to enhance its potential for publication.
In the Introduction and rationale, it would be necessary to provide a stronger theoretical justification for the use of the BFI-10, particularly in comparison with other versions that have already been validated in the region, such as the 44-item version. Although the manuscript cites studies conducted in countries such as France, Brazil, and India, these references may not be sufficient to establish cross-cultural confidence, especially considering that the BFI-10 has not consistently produced the expected psychometric results across different settings.
It would also be useful to expand the methodological justification for the restrictions imposed in the confirmatory factor analysis, given the structure of the dimensions. In addition, the appropriateness of using Cronbach's alpha and McDonald's omega for two-item dimensions should be more clearly justified. The authors may consider reporting average inter-item correlations as complementary evidence of reliability.
Finally, the criterion validity results do not appear to be particularly encouraging, which contrasts with the statement in the Discussion that the instrument "can be considered a valid and reliable tool for assessing personality traits in Peruvian medical students." This conclusion should be moderated. The authors should also discuss more critically the finding related to the Openness to Experience dimension, whose reliability was below the recommended threshold.

Flores-Cohaila, J. A., Miranda-Chavez, B., Huarcaya-Victoria, J., Copaja-Corzo, C., Montesinos, S. Y., & Estela-Fernandez, C. A. (2026). Evidence of validity of the 10-item Big Five Inventory (BFI-10) in Peruvian medical students. Interacciones, 12, e542. https://doi.org/10.24016/2026.v12.542

RESPONSE LETTER

Title: Evidence of validity of the 10-item Big Five Inventory (BFI-10) in a medical student population **Journal:** Interacciones

Dear Editor and Reviewers,

We sincerely thank the Editor and the three reviewers for their careful, constructive, and expert evaluation of our manuscript. The comments were unusually substantive and have materially improved the work.

One reviewer observation deserves to be highlighted up front because it bears on several comments. Reviewers D and E correctly identified that **the factor labels in Table 3 were misassigned**. We are grateful for this catch. Upon verification we confirmed the error and **re-ran the entire analysis from the raw dataset**. We report the outcome transparently: the misassignment was confined to the *labels*; the underlying model, fit indices, and loadings reproduce. After relabeling, two substantive consequences emerged, both of which we have incorporated:

1. The criterion-validity test against psychiatric history had inadvertently been computed on the wrong dimension. With the corrected labels, **Neuroticism is significantly associated with prior mental-health care** (Welch t = 3.00, p = 0.004, Cohen's d = 0.55) — a theoretically expected association that the original (mislabeled) analysis had missed.
2. When academic performance is analyzed as a continuous variable (as Reviewer E recommends), **no Big Five dimension correlates significantly with GPA**; the previously reported r = 0.13 did not survive.

Accordingly, we have moderated all claims about criterion validity, corrected the tables, and re-verified that every downstream interpretation now refers to the correctly labeled factors. Below we respond to each comment point by point. Changes are marked in the revised manuscript.

Reviewer D

RD.1 — Critical factor mislabeling in Table 3. > *"The item content in Table 3 does not match the factor labels… BF5–BF6 to Conscientiousness; BF7–BF8 to Neuroticism; BF9–BF10 to Openness… The authors must correct the table and verify that all interpretations… refer to the correctly labeled factors."*

We thank the reviewer and fully accept this comment. The reviewer is correct. We re-estimated the model from the raw data and corrected the factor assignment throughout: **Openness = BF5, BF6** (original, inventive); **Conscientiousness = BF7, BF8** (efficient, hard-working); **Neuroticism = BF9, BF10** (gets tense, nervous). Table 3 has been rebuilt with the correct labels, and we verified that the CFA, measurement-invariance, reliability, and criterion-validity sections now all refer to the correctly labeled factors. The corrected standardized loadings are: Extraversion 0.89/0.85; Agreeableness 0.80/0.73; Openness 0.72/0.71; Conscientiousness 0.79/0.72; Neuroticism 0.79/0.79 (all $p < 0.001$). See revised Table 3 and Results [p. 12].

RD.2 — Implausible inter-factor correlations / equality constraint. > *"…values of 0.84 (Conscientiousness–Openness) and 0.75 (Agreeableness–Conscientiousness)… A likely contributor is the equality constraint… The authors should acknowledge this explicitly and discuss whether these correlations… are an artifact of the constrained model."*

We thank the reviewer for this insightful comment, which we accept. With corrected labels, the high correlations are Agreeableness–Openness (0.85) and Openness–Conscientiousness (0.84). To address whether these are an artifact of the equality constraint, we re-estimated the model **without** the constraint. The high correlations persisted (0.84 and 0.82), indicating they are not an artifact of the constraint but reflect genuinely low discriminant separation among Agreeableness, Openness, and Conscientiousness when each is measured by only two items — a known limitation of ultra-short inventories. Notably, the unconstrained model produced an improper (Heywood) solution for Neuroticism, confirming that the equality constraint is required for identification of these two-indicator factors rather than being a matter of convenience. We now state this explicitly in the Methods and discuss the discriminant-validity limitation in the Discussion [p. 5-6].

RD.3 — Borderline configural fit for sex undermines invariance. > *"The configural model for gender… CFI = 0.930 and RMSEA = 0.092… Scalar invariance derived from a poorly-fitting baseline model cannot be considered robust."*

We thank the reviewer for this insightful comment and we accept this comment. The re-analysis confirms a marginal configural baseline for sex (RMSEA ≈ 0.098). We have substantially qualified the corresponding claims: we now state that sex-based invariance should be interpreted with caution because the configural baseline exceeds the RMSEA threshold, and we have removed any strong claim of robust scalar invariance

across sex. The invariance by academic phase (configural RMSEA = 0.077) is retained with appropriate caveats. See Results and Limitations [p. 5].

RD.4 — Criterion validity overstated. > *"The only significant external validity finding is r = 0.13… These results constitute very weak criterion validity evidence. The abstract and conclusions should be revised…"*

We thank the reviewer for this comment, we have revised the manuscript accordingly. After re-analysis with continuous GPA and corrected labels, no dimension correlates significantly with GPA. We now state plainly that the BFI-10 shows **adequate structural validity but insufficient and preliminary criterion-validity evidence**. We also report the one association that did emerge after correction (Neuroticism–psychiatric history, p = 0.004, d = 0.55) as preliminary supporting evidence, while being explicit about the GPA null result. The Abstract, Results, Discussion, and Conclusions have been revised to reflect this. See [Abstract: p. 1; Results; Discussion; Conclusions: p. 4-5].

RD.5 — Wrong reference for the BFI-10 ultra-short version. > *"'the ultra-short version reported by Costa [ref. 8]' should reference Rammstedt & John (2007) [ref. 6]… Reference 8 is a Brazilian validation study."*

We thank the reviewer for this insightful comment and have corrected this section. In the Variables section, the ultra-short BFI-10 is now attributed to **Rammstedt & John (2007)** [ref. 6]; the Costa et al. (2023) Brazilian study is cited where appropriate as a comparative validation. See Methods, Variables [p. 3].

RD.6 — "responsibility" → "conscientiousness". > *"The term 'responsibility'… should be 'conscientiousness' throughout."*

We thank the reviewer for this comment, we have corrected the terminology throughout the Results and Discussion.

RD.7 — Table 4 headers in Spanish. > *"Table 4 headers and labels are in Spanish; they should be translated to English."*

Table 4 headers, labels, and footnotes have been fully translated to English. See revised Table 4 [p. 13].

Reviewer E

RE.1 — Strengthen rationale with Latin American examples and short-instrument failures; expand BFI-10 validity evidence. > *"…include examples of evidence derived from personality assessments conducted in Latin America, as well as examples of brief instruments that have failed to demonstrate robust psychometric properties… the discussion of the available validity and reliability evidence for the BFI-10 should be expanded."*

We thank the reviewer for this insightful comment. We have expanded the Introduction (paragraph 2) with Latin American personality-assessment evidence (including the Peruvian BFI validation by Dominguez-Lara et al. and the Argentine adaptation by Genise et al.) and with examples where ultra-short instruments did not achieve robust structure (e.g., the BFI-10 in Indian adolescents and in the Brazilian general population). We also expanded the comparison of existing BFI-10 validity/reliability evidence against our findings in the Discussion. See Introduction [p. 2] and Discussion [p. 5-6].

RE.2 — Recent references justifying two items per trait. > *"…support, with more recent references, the rationale for considering two items sufficient… several studies recommend a minimum number of items."*

We thank the reviewer for this comment. We added a justification in the Methods and Discussion acknowledging that two-item factors are a deliberate trade-off of ultra-short inventories, supported by recent psychometric literature, and we explicitly note the reliability cost (see RF.4). See Methods, Data analysis [p. 3-4].

RE.3 — Table 1 "population" → "sample"; APA formatting. > *"…the term 'sample' should be used instead of 'population'… the tables are not formatted according to APA guidelines."*

We thank the reviewer for this comment. The Table 1 title now reads "sample," and all tables have been reformatted to APA 7th-edition style (titles, notes, alignment, no vertical rules). See revised Tables 1–4.

RE.4 — Identify which items belong to each dimension; verify high within / low between correlations. > *"…clearly identify which items correspond to each dimension in order to verify that the correlations between the two items within each dimension are sufficiently high, while correlations with items from other dimensions remain low."*

We thank the reviewer for this insightful comment. The corrected Table 3 now explicitly maps each item to its dimension, and we added the within-dimension inter-item correlations (Extraversion 0.75, Agreeableness 0.58, Openness 0.51, Conscientiousness 0.58, Neuroticism 0.62). We also note in the Results that Neuroticism items (BF9, BF10) correlate strongly with each other and near-zero with all other items, supporting their separation. See Results [p. 4] and Table 3.

RE.5 — Table 3 items 7–10 misclassified (openness ↔ neuroticism). > *"In Table 3, Items 7–10 appear to be misclassified, as the items assigned to openness actually correspond to neuroticism, and vice versa."*

We thank the reviewer; this is correct and converges with RD.1. The factor assignment has been fully corrected and all analyses re-verified (see our response to RD.1). See revised Table 3.

RE.6 — Invariance across sex, not gender. > *"Measurement invariance was evaluated across sex, not gender; consequently, the terminology… should be modified."*

We thank the reviewer for this comment. We have corrected the use of use "sex" consistently in the text, Table 4, and footnotes.

RE.7 — Table 4 categories in Spanish. > *"The categories presented in Table 4 are written in Spanish… should therefore be translated, along with the corresponding table footnotes."*

We thank the reviewer for this comment. Table 4 categories and footnotes are now in English (see also RD.7).

RE.8 — Dichotomous treatment indicator insufficient; add convergent/divergent variables. > *"The dichotomous indicator of prior psychological or psychiatric treatment may be insufficient… More detailed measures… are necessary… it is recommended that convergent and divergent validity evidence be examined using additional variables."*

We thank the reviewer for this insightful comment. We acknowledge the dichotomous indicator as a limitation and have strengthened the Limitations section accordingly (no timing or symptom severity, no validated comparator instruments). Within the available data, we now report the corrected Neuroticism–psychiatric history association ($p = 0.004$, $d = 0.55$) as preliminary convergent evidence, and we explicitly flag the need for validated comparators (e.g., PHQ-9, GAD-7, a full BFI-44) in future work. See Limitations [p. 5] and Discussion [p. 6].

RE.9 — Use continuous GPA, not categorized. > *"The categorization of academic average scores into intervals… is not recommended when exact continuous scores are available… loss of information."*

We thank the reviewer for this comment. We re-analyzed criterion validity using **continuous GPA** with Spearman correlations. No dimension reached significance (Conscientiousness $\rho = 0.10$, $p = 0.13$; all others $|\rho| \leq 0.07$). The categorization was removed from the analysis and Results updated accordingly. See Methods [p. 3] and Results [p. 4-5].

RE.10 — Vancouver → APA referencing. > *"The manuscript is currently formatted using a Vancouver-style referencing system rather than the APA format required by the journal guidelines."*

We thank the reviewer for this comment. Both in-text citations and the reference list have been converted to APA 7th-edition style throughout. See References and all in-text citations.

Reviewer F

RF.1 — Stronger theoretical justification for BFI-10 vs. validated longer versions (e.g., 44-item); cross-cultural caution. > *"…provide a stronger theoretical justification for the use of the BFI-10, particularly in comparison with… the 44-item version… the BFI-10 has not consistently produced the expected psychometric results across different settings."*

We thank the reviewer for this comment. We expanded the Introduction to justify the BFI-10 specifically (feasibility in large-scale, time-constrained, multi-site research where the BFI-44 is impractical) while explicitly acknowledging its inconsistent cross-cultural performance, and we frame our study as testing whether the structure holds in a previously unrepresented population. See Introduction [p. 2] and Discussion [p. 3-4].

RF.2 — Justify CFA restrictions; appropriateness of alpha/omega for two-item dimensions; report average inter-item correlations. > *"…expand the methodological justification for the restrictions… the appropriateness of using Cronbach's alpha and McDonald's omega for two-item dimensions should be more clearly justified… consider reporting average inter-item correlations…"*

We thank the reviewer for this comment. The Methods now justify the equality constraint as required for identification of two-indicator factors (see RD.2, including the Heywood result without it). We acknowledge that alpha/omega are limited for two-item scales and now **report the inter-item correlation for each dimension** as complementary reliability evidence (Table 3). See Methods [p. 3] and Table 3.

RF.3 — Moderate the "valid and reliable tool" claim. > *"…the criterion validity results do not appear to be particularly encouraging, which contrasts with the statement… that the instrument 'can be considered a valid and reliable tool…'. This conclusion should be moderated."*

We thank the reviewer for this comment, this converges with RD.4. We have removed the unqualified "valid and reliable tool" statement. The Discussion and Conclusions now state that the BFI-10 shows adequate structural validity and measurement invariance (with the sex caveat), but that criterion-validity evidence remains preliminary and the instrument is suitable for group-level research rather than individual or selection decisions. See Discussion and Conclusions [p. 5-6].

RF.4 — Critically discuss Openness reliability below threshold. > *"The authors should also discuss more critically the finding related to the Openness to Experience dimension, whose reliability was below the recommended threshold."*
We thank the reviewer for this comment. After correcting the factor labels, **Openness (BF5, BF6) is indeed the dimension with reliability below 0.70 (ω = 0.68, α = 0.68)**. We now discuss this critically in the Discussion and Limitations: the two Openness items show modest inter-item correlation (0.51), individual-level Openness scores should be interpreted with caution, and we recommend additional Openness items in future applications. See Discussion [p. 6] and Limitations [p. 5].

We are grateful for the time and expertise the reviewers and Editor invested in this manuscript. We believe the revised version is substantially stronger and we hope it now meets the standards of *Interacciones*.
Sincerely, On behalf of all authors, Cesar Copaja-Corzo (corresponding author)