LETTER OF REVIEWERS

Reviewer B:
Recommendation: Revisions Required

Relevance: High
Novelty: Moderated
Presentation and writing: High

Comments for authors:

General Evaluation: The study presents a relevant design and addresses a theoretically significant instrument. However, methodological improvements are needed in the psychometric justification of the instrument, the reporting of data analyses, and the consistency between the sample scope and the study's conclusions.

Title: The title is appropriate, as it accurately describes the scope of the study and clearly indicates that the sample consists of university students. However, for this same reason, it is important to emphasize that the sample is not representative of the general population. Therefore, it is recommended to include a reference to the exploratory nature of the study in the title (e.g., "Preliminary study...") in order to convey that this is a first step that should later be replicated in a representative sample of the Peruvian population.

Abstract: The abstract is well-structured and clearly presents the main components of the study. However, it lacks information on which Spanish version of the instrument was used, which is relevant from both a methodological and psychometric perspective.

Introduction: The introduction is appropriate and adequately contextualizes the construct of interest. However, a more detailed psychometric description of the GQ-6 instrument is missing. Specifically, the authors should: describe the internal structure or dimensionality of the original version of the instrument, as proposed by McCullough et al. (2002); and report previous studies that have evaluated the factorial structure of the GQ-6 using confirmatory factor analysis (CFA), indicating whether this structure has been confirmed or questioned in prior research.
The discussion of issues related to item 6 is appropriately presented and constitutes a relevant contribution. However, it would be useful to emphasize that many of these issues arise from models that do not fit well in CFA and only achieve acceptable fit after model respecification.
Finally, the formulation of the study's objective should explicitly state that the sample consists of university students, to maintain consistency with the title and the actual scope of the study.

Method - Participants: This section states that a non-probabilistic convenience sampling method was used, selecting participants based on accessibility. This information clearly defines the scope of the study and aligns with the title. However, this characteristic of the sample should also be explicitly reflected in the study's objectives and in the limitations section of the discussion. For this reason, the recommendation to use the term "preliminary study" in the title is reinforced to maintain methodological coherence.

Method – Instruments: This section describes the five instruments used in the study. In all cases, the following should be specified: which Spanish version is being used and a brief report of its main psychometric properties. Regarding reliability, it is not sufficient to report only the values obtained in the original version of the instrument. It is also necessary to report the reliability obtained with the sample of the present study (induced reliability), as recommended by current psychometric guidelines. In this regard, it is suggested to follow the recommendations of the American Psychological Association for reporting empirical results, as outlined by Sánchez-Meca, López-Pina, and López-López (2008).

Method - Data Analysis: The authors propose several fit indices to evaluate the model, but they apply cut-off points that do not align with current methodological recommendations. Hair et al. (2009, 2010) recommend establishing fit criteria based on the number of items and the sample size. In this case, with a six-item model ($m < 12$) and a sample of $N = 440$ ($N > 250$), the following should be considered indicators of good fit: CFI and TLI ≥ .95, RMSEA < .07, and SRMR ≤ .08, acknowledging that the χ^2 statistic may be significant even for correctly specified models.

Additionally, the recommendations of Schreiber et al. (2006) can be incorporated for complementary indices (e.g., $\chi^2/df \leq 2–3$, CFI ≥ .95, TLI ≈ .96, SRMR ≤ .08, WRMR < .90, and RMSEA < .06) for categorical data. If the items do not present relevant skewness and kurtosis issues, and considering that a Likert scale with more than five categories is used, it would also be appropriate to estimate the model using an estimator for continuous variables. In that case, the following criteria can be applied: CFI ≥ .95, TLI ≈ .95, SRMR ≤ .08, WRMR < .90, and RMSEA < .08.

Results: A table reporting descriptive statistics for the six items is missing. This table should include means, standard deviations, skewness and kurtosis indices, discrimination indices, and reliability analysis if each item were removed. This information is particularly relevant for scales with seven response options and would provide stronger justification for the choice of estimator in the CFA.

Discussion: Table 2 shows that the GQ-5 model exhibits a very good fit in most indices, except for RMSEA. Before proceeding with model respecification by correlating the error terms of items 4 and 5, it is necessary to report the modification indices that support this decision. The correlation of residual errors is acceptable only when there is a clear substantive justification (e.g., very similar wording, overlapping content, double content, or method effects). Otherwise, it is a post hoc respecification that may lead to model overfitting.

Psychometric literature is clear on this point. As Brown (2015) states: “As with any type of parameter specification in CFA, correlated errors must be supported by a substantive rationale and should not be freely estimated simply to improve model fit” (p. 161).

Therefore, it is recommended to explicitly justify the decision to allow residual correlations and to discuss their implications for the construct validity of the instrument.

RESPONSE LETTER

I thank the reviewer for the careful evaluation of this manuscript. The comments provided were helpful in improving its clarity and rigor. All revisions have been incorporated into the revised manuscript and are highlighted in gray for ease of identification.

Title

1. The title is appropriate, as it accurately describes the scope of the study and clearly indicates that the sample consists of university students. However, for this same reason, it is important to emphasize that the sample is not representative of the general population. Therefore, it is recommended to include a reference to the exploratory nature of the study in the title (e.g., "Preliminary study...") in order to convey that this is a first step that should later be replicated in a representative sample of the Peruvian population.

Thank you for this suggestion. To clarify the scope of the study and avoid overgeneralization beyond the university population, the title has been revised to explicitly indicate the preliminary nature of the psychometric evidence.

Abstract

2. The abstract is well-structured and clearly presents the main components of the study. However, it lacks information on which Spanish version of the instrument was used, which is relevant from both a methodological and psychometric perspective.

Thank you for this observation. The abstract has been revised to explicitly indicate that the Spanish version of the Gratitude Questionnaire adapted by Quezada-Berumen et al. (2023) was used.

Introduction

3. The introduction is appropriate and adequately contextualizes the construct of interest. However, a more detailed psychometric description of the GQ-6 instrument is missing. Specifically, the authors should: describe the internal structure or dimensionality of the original version of the instrument, as proposed by McCullough et al. (2002); and report previous studies that have evaluated the factorial structure of the GQ-6 using confirmatory factor analysis (CFA), indicating whether this structure has been confirmed or questioned in prior research.

Thank you for this valuable comment. This issue has been addressed in the third paragraph of the Introduction, where the original unidimensional structure of the GQ-6 proposed by McCullough et al. (2002) is now explicitly described.

In addition, prior studies using confirmatory factor analysis (CFA) across different cultural contexts are summarized, including evidence that has questioned the adequacy of the six-item model due to poor global fit indices.

4. The discussion of issues related to item 6 is appropriately presented and constitutes a relevant contribution. However, it would be useful to emphasize that many of these issues arise from models that do not fit well in CFA and only achieve acceptable fit after model respecification.

Thank you for this insightful comment. This suggestion has been addressed in the fourth paragraph of the Introduction, where it is now emphasized that problems related to Item 6 have been consistently reported in studies showing inadequate CFA model fit, and that improved fit is typically achieved after addressing these fit-related issues through item removal, resulting in the GQ-5.

5. Finally, the formulation of the study's objective should explicitly state that the sample consists of university students, to maintain consistency with the title and the actual scope of the study.

In the last paragraph of the Introduction, the study objective has been revised to explicitly specify that the sample consists of Peruvian university students, ensuring full consistency with the title and the scope of the study.

Method

Participants

6. This section states that a non-probabilistic convenience sampling method was used, selecting participants based on accessibility. This information clearly defines the scope of the study and aligns with the title. However, this characteristic of the sample should also be explicitly reflected in the study's

objectives and in the limitations section of the discussion. For this reason, the recommendation to use the term “preliminary study” in the title is reinforced to maintain methodological coherence.
Thank you for this valuable comment. In response, the title has been revised to explicitly reflect the preliminary nature of the study.
In addition, the Participants section was clarified to indicate that the use of a non-probabilistic convenience sample is consistent with the preliminary and exploratory scope of the research.

Instruments

7. This section describes the five instruments used in the study. In all cases, the following should be specified: which Spanish version is being used and a brief report of its main psychometric properties.
The Instruments section has been revised to explicitly indicate the Spanish version used for the GQ-6, specifying the adaptation by Quezada-Berumen et al. (2023).

8. Regarding reliability, it is not sufficient to report only the values obtained in the original version of the instrument. It is also necessary to report the reliability obtained with the sample of the present study (induced reliability), as recommended by current psychometric guidelines. In this regard, it is suggested to follow the recommendations of the American Psychological Association for reporting empirical results, as outlined by Sánchez-Meca, López-Pina, and López-López (2008).
I appreciate this comment and would like to clarify that the reliability of the GQ-6 and GQ-5 is empirically examined and reported in the Results section, as the primary aim of the study is to evaluate the psychometric properties of these instruments.
For the remaining instruments, reliability estimates obtained in the present sample were already reported in the Measures section.

Data Analysis

9. The authors propose several fit indices to evaluate the model, but they apply cut-off points that do not align with current methodological recommendations. Hair et al. (2009, 2010) recommend establishing fit criteria based on the number of items and the sample size. In this case, with a six-item model ($m < 12$) and a sample of $N = 440$ ($N > 250$), the following should be considered indicators of good fit: CFI and TLI ≥ .95, RMSEA < .07, and SRMR ≤ .08, acknowledging that the χ^2 statistic may be significant even for correctly specified models.
Thank you for this methodological observation. In response, the model fit criteria have been revised to align with current recommendations that consider model complexity and sample size. Specifically, more stringent cut-off values appropriate for short scales and large samples were adopted (CFI/TLI ≥ .95, RMSEA < .07, SRMR ≤ .08), following Hair et al. (2009, 2010). The Data Analysis section has been updated accordingly, and the χ^2 statistic is now explicitly interpreted with caution.

10. Additionally, the recommendations of Schreiber et al. (2006) can be incorporated for complementary indices (e.g., $\chi^2/df \leq 2–3$, CFI ≥ .95, TLI ≈ .96, SRMR ≤ .08, WRMR < .90, and RMSEA < .06) for categorical data. If the items do not present relevant skewness and kurtosis issues, and considering that a Likert scale with more than five categories is used, it would also be appropriate to estimate the model using an estimator for continuous variables. In that case, the following criteria can be applied: CFI ≥ .95, TLI ≈ .95, SRMR ≤ .08, WRMR < .90, and RMSEA < .08.
I appreciate this comment. However, given the ordinal nature of the data and the use of the WLSMV estimator, model evaluation focused on a parsimonious set of fit indices that are most consistently recommended and interpretable under this estimation framework (i.e., RMSEA, CFI, TLI, and SRMR). Indices such as χ^2/df and WRMR were not emphasized due to their sensitivity to sample size and limited interpretability under WLSMV.

Results

11. A table reporting descriptive statistics for the six items is missing. This table should include means, standard deviations, skewness and kurtosis indices, discrimination indices, and reliability analysis if each item were removed. This information is particularly relevant for scales with seven response options and would provide stronger justification for the choice of estimator in the CFA.
Thank you for this suggestion. Table 1 in the Results section has been expanded to include means, standard deviations, skewness, kurtosis, corrected item–total correlations as discrimination indices, and reliability if item deleted (McDonald’s omega).

12. Table 2 shows that the GQ-5 model exhibits a very good fit in most indices, except for RMSEA. Before proceeding with model respecification by correlating the error terms of items 4 and 5, it is necessary to report the modification indices that support this decision.
Thank you for this observation. The modification index supporting the residual correlation between Items 4 and 5 has now been explicitly reported in the third paragraph of the Results section (MI = 14.96), and the rationale for freely estimating this covariance is clarified in the revised Discussion.

Discussion
13. The correlation of residual errors is acceptable only when there is a clear substantive justification (e.g., very similar wording, overlapping content, double content, or method effects). Otherwise, it is a post hoc respecification that may lead to model overfitting.
Psychometric literature is clear on this point. As Brown (2015) states: “As with any type of parameter specification in CFA, correlated errors must be supported by a substantive rationale and should not be freely estimated simply to improve model fit” (p. 161).
Therefore, it is recommended to explicitly justify the decision to allow residual correlations and to discuss their implications for the construct validity of the instrument.
Thank you for this important comment. The manuscript has been revised to explicitly provide a substantive justification for the correlated residuals between Items 4 and 5, based on their overlapping content.
Also, the implications of this specification for the unidimensional structure and construct validity of the GQ-5 are now explicitly discussed in the third paragraph of the Discussion section.