



Enhancing Rigor in Nursing Research: A Critical Analysis of Reliability and Validity Reporting Trends

Khalil Tazik¹, Mahmood Maniati^{1,*}

¹ School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

*Corresponding Author: School of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. Email: maniatim@yahoo.com

Received: 1 June, 2024; Revised: 24 September, 2024; Accepted: 29 September, 2024

Abstract

Context: In recent years, there has been a growing emphasis on ensuring the accuracy and reliability of research findings. However, the reporting of reliability and validity measures in nursing research articles (RAs) remains underexplored.

Objectives: This study aimed to investigate the current trends in reliability and validity reporting within nursing RAs and explore the implications of these practices.

Methods: The study analyzed 42 RAs published between 2021 and 2022 from seven key nursing journals. The inclusion criteria were defined to incorporate empirical studies that employed tests and/or questionnaires. To maintain consistency and homogeneity, we exclusively included studies that relied on data collected through these methods. Title and abstract screening were performed by two independent reviewers, followed by a full-text review.

Results: The study revealed a concerning lack of attention to reliability and validity reporting. A significant portion of the reviewed articles (19.05%) did not explicitly report reliability measures, and an even larger proportion (38.10%) failed to report validity measures.

Conclusions: Our findings highlight the need for a significant change in how nursing research reports reliability and validity measures. By improving the clarity of these reports, researchers can enhance the trustworthiness of their findings and ensure that they can be applied more broadly in nursing practice.

Keywords: Nursing, Reliability, Validity, Construct Validity

1. Context

The consensus among researchers is unequivocal, emphasizing the paramount importance of legitimizing and rigorously substantiating research findings (1). In adhering to this principle, scientific research must steer clear of haphazardness or lack of rigor (1). The continual assessment of the quality and authenticity of research findings is deemed essential, with particular attention to the information acquired through measurement instruments and the subsequent conclusions and interpretations drawn from this data (2).

Recent studies published in the past few years have once again emphasized the crucial significance of guaranteeing the accuracy and dependability of research outcomes. Stachl et al. (3) argue that, in an era marked by rapidly evolving research methodologies, the

validity of data becomes even more central to the credibility of scientific inquiry. These findings emphasize the necessity for researchers to modify their validation procedures to tackle the intricacies that arise from the emergence of new technologies and methodologies. Moreover, the work of Klein (4) delves into the evolving landscape of research quality assurance. Indeed, in an era of increasing interdisciplinary research, ensuring the reliability of findings across diverse domains is essential. The obtained results indicate that interdisciplinary research encounters distinct obstacles in upholding both internal and external credibility, highlighting the need for a sophisticated validation process (4).

Furthermore, the concept of "credibility rules" in research, as discussed by Gethmann et al. (5), has been expanded upon by Rieh and Danielson (6). Their most

recent research highlights the importance of establishing a thorough framework for evaluating credibility, which includes not just methodological precision but also transparency and ethical factors. This acknowledgment considers the evolving needs of both the scientific community and the broader public (6).

In the domain of medical sciences and nursing, recent literature by Cruz Rivera et al. (7) explores the challenges and opportunities in ensuring the validity of patient-reported outcomes in clinical research. Their study emphasizes the importance of adopting standardized and validated measures to capture meaningful patient perspectives, ultimately contributing to more robust clinical decision-making.

In combining the most recent advancements spanning various fields, it is apparent that the quest for dependable and valid research results is a continuous and multidisciplinary effort. This necessitates continual adaptation to the evolving landscape of research methodologies, not only within linguistics but also in medical sciences and nursing, highlighting the importance of a comprehensive approach to validation and reliability in diverse fields (7).

1.1. Reliability in Academic Research: A Contemporary Perspective

Reliability, defined by Dörnyie (1) as "the extent to which our measurement instruments and procedures produce consistent results in a given population in different circumstances" (p. 50), remains a fundamental criterion in research quality assessment. Recent scholarship has further illuminated the multifaceted nature of reliability and its implications across diverse disciplines (1).

Variations in test methodologies, assessors, examinees, and the test itself (8) may introduce inconsistencies, resulting in unreliable findings. Researchers using tests and surveys for data gathering or working with multiple assessors or observers expect reliable results, irrespective of fluctuations in test elements or the number of assessors and observers (9). Recent literature by Elliot et al. (10) explores the evolving landscape of reliability assessment in the context of advanced research methodologies. In the current era characterized by technological progress and multidisciplinary investigation, it is contended that the critical task of tackling the obstacles presented by novel approaches is essential for safeguarding the

dependability of results. Their work highlights the need for researchers to adapt traditional reliability measures to meet the demands of contemporary research practices (10).

It is crucial to clarify that reliability constitutes a psychometric attribute of test scores derived from a specific respondent group, as opposed to being an intrinsic quality of the instrument itself. Onwuegbuzie and Leech (11) emphasize the widespread neglect of reliability estimates in quantitative research, pointing out a common oversight in research practices. Dörnyie (1) attributes this omission to a misdirected assumption that reliability relates to the characteristics of tests and research instruments rather than the scores they produce. Consequently, Bachman (12) advocates for researchers to disclose reliability estimates for the test scores associated with the instruments they utilize.

Weir (13) underscores the crucial significance of understanding the reliability of a test, asserting that without knowing the reliability of a test, there is no way to know how consistent the results are. Bachman (8) presents two methodologies for gauging internal consistency when communicating the reliability of test scores: (1) an estimate based on the correlation between two sets of scores; and (2) estimates based on ratios of the variances of halves or items of the test to the total test score variance. Eisinga et al. (14) agree with the assertion, highlighting the correlation coefficient as the predominant measure used to communicate reliability, functioning as a quantitative depiction of the extent of associations among variables.

In the domain of reliability coefficients, diverse statistical approaches are utilized in different studies. For instances of inter-rater or intra-rater reliability, test-retest, and parallel form correlation coefficients are applied. Meanwhile, for assessing the internal consistency of items, researchers often utilize correlation, Spearman-Brown, Cronbach alpha (coefficient alpha), Kuder-Richardson 20 & 21. Cronbach alpha and Kuder-Richardson 20 & 21 are notably common methods for communicating internal consistency (15).

Acknowledging the crucial role of score reliability, Vacha-Haase et al. (15) emphasize its direct influence on test outcomes and interpretations. Despite the importance of reliability, there is no explicitly defined model for reporting validity and reliability across various editions of the APA Publication Manual. The

manual encourages authors to furnish ample information regarding the validity and reliability of their findings (16).

1.2. Validity in Quantitative Research: A Contemporary Examination

Dörnyie (1) outlined reliability as a relatively straightforward concept in quantitative research, but when addressing validity, he identified two parallel systems: Construct validity (measurement validity) and its components, and the internal/external validity dichotomy (research validity). Introducing a layer of complexity, Dörnyie (1) included reliability as the third dimension in discussions related to quantitative quality standards. Research validity, encompassing external and internal validity, concerns the overall soundness of the research process. Internal validity pertains to the correlation between the variables under study and the results obtained, whereas external validity concerns the extent to which research outcomes can be applied to broader contexts. Measurement validity, emphasized by Dörnyie (1), revolves around 'the meaningfulness and appropriateness of the interpretation of the various test scores or other assessment procedure outcomes' (p. 50). This can be seen as a unified concept (16) expressed in terms of construct validity, or it can be broken down into content, criterion, and construct validity. The main concern lies in interpreting test scores accurately, not just ensuring the scores themselves are valid (8). Chinni and Hubley (17) offered insights into validation practices, characterizing them as tools researchers employ to construct arguments and justifications for test score inferences or explanations (p. 36). Zumbo (18) reinforced the significance of validity by asserting that without it, inferences drawn from research instruments are rendered meaningless.

In summary, the understanding of validity in quantitative research has evolved to encompass not only traditional dimensions but also contemporary challenges posed by technological advancements and interdisciplinary research. Researchers must adapt validation practices to ensure the meaningfulness of inferences, considering the specific contexts and domains in which their research is applied. This research examines how reliability and validity are reported in current nursing studies, highlighting the problems caused when these essential aspects are ignored.

2. Evidence Acquisition

2.1. Methodology: Constructing and Analyzing the Corpus

2.1.1. Corpus Selection

The primary dataset for this study comprised 42 Research Articles (RAs) published within the timeframe of 2021 to 2022 (Appendix 1). Seven key journals in the field of nursing were chosen for this study: Clinical Simulation in Nursing, European Journal of Oncology Nursing, International Journal of Nursing Studies, Journal of Cardiovascular Nursing, Journal of Nursing Scholarship, Journal of Professional Nursing, Nurse Education Today, and Nursing Outlook. These nursing journals were selected due to their recognized standing in the field, ensuring the inclusion of high-impact and influential research. This strategic choice enhances the generalizability of findings to a broader academic audience within the domain of nursing research. The inclusion criteria were defined to incorporate empirical studies employing tests and/or questionnaires. To maintain consistency (homogeneity), we only included studies that relied on data collected through tests or questionnaires. This meant excluding studies that were non-empirical or used different methods like interviews or observations.

2.1.2. Analysis

To ensure a robust and comprehensive analysis, a team of four analysts, consisting of the primary researchers and two colleagues, collaborated in the coding process. Each analyst independently evaluated the coded papers, fostering a diverse and multi-perspective examination of the corpus. To validate the consistency and reliability of the individual evaluations, a dedicated session was conducted to compare and reconcile any disparities in the assessments. Resolving discrepancies involved thorough discussions, with the ultimate decision-making responsibility resting with the two main authors.

To systematically analyze validity and reliability reports within the corpus, a comprehensive coding sheet was developed. This coding sheet served as the primary instrument for capturing and categorizing the diverse validity and reliability evidence presented in each research article. Each piece of evidence was uniquely identified and linked to the respective paper,

facilitating a nuanced examination of reporting practices. The coding sheet not only captured the validity sources but also incorporated authors' justifications for employing specific types of validity measures. This qualitative dimension added depth to the analysis, shedding light on the rationale behind the selection of particular validity assessments. The inclusion of author justifications aimed to provide a richer understanding of the thought processes influencing validity reporting in the analyzed studies.

Drawing from established frameworks in the field, the coding of validity evidence incorporated a range of sources, aligning with the taxonomy outlined by Zumbo et al. (19). This study examined a range of validity types commonly used in research, such as face validity, content validity, construct validity, predictive validity, concurrent validity, convergent validity, discriminant validity, response processes validity, consequences validity, reliability, internal structure validity, among others (19). This comprehensive set of validity sources was coded based on explicit or implicit mentions in the analyzed studies. In cases where the literature implicitly suggested the application of certain validity measures, the coding process involved deductive reasoning. For example, if a research article indicated the use of factor analysis for questionnaire validation, it was deduced that construct validity was implicitly assessed. The coding sheet documented both explicit mentions and inferred validity sources, contributing to a thorough and nuanced analysis.

For reliability, the coding encompassed various dimensions, including internal consistency, parallel form, test-retest, and inter-rater evidence. Each article was categorized based on the types of reliability evidence reported, contributing to a comprehensive overview of reliability reporting practices.

We grouped the articles based on how they addressed the accuracy of their measurements. Some articles only considered validity, others only considered reliability, and some addressed both, while some didn't mention either. This classification schema facilitated the identification of prevalent reporting patterns within the corpus (Table 1).

3. Results

To start with, a classification of the validity and reliability methods used in nursing RAs is provided. By classifying the studies, we can observe the different

approaches researchers employed to assess the quality of their data (validity and reliability) and the statistical tools they utilized for analysis.

Table 2 presents a comprehensive overview of the psychometric properties employed in nursing RAs for instrument validation and reliability assessment. Construct validity, examined in 45.24% of studies, primarily utilized Factor Analysis, emphasizing a focus on understanding the underlying structure of measured constructs. Internal structure assessment, conducted in 28.57% of studies, employed Confirmatory and Exploratory Factor Analysis, indicating a dedication to exploring relationships among variables. Content validity, vital for instrument relevance, was addressed in 19.05% of studies through iterative expert assessments, literature searches, and pilot studies. Face validity, ensuring instrument appropriateness, was assessed in 16.67% of studies using expert panels and pilot tests. Convergent validity (14.29%) and psychometric evaluation (11.90%) were explored through Confirmatory Factor Analysis and discriminant validity testing, revealing a comprehensive approach to validity assessment. To evaluate concurrent validity, discriminant validity, and criterion validity (all at 7.14%), researchers relied on Pearson's correlation coefficient. Predictive validity (4.76%) utilized Pearson's correlation coefficient, and two studies (4.76%) implicitly investigated response processes. Two studies (4.76%) did not explicitly mention validation processes, and another two (4.76%) relied on literature-based validation. Consequential validity (2.38%) utilized Structural Equation Modeling Modeling to explore the impact of instrument use. Remarkably, 38.10% of studies did not specify the validity type assessed, highlighting a need for improved clarity.

For reliability, 50% of studies assessed internal consistency using Cronbach's alpha and reliability testing. Researchers used correlation coefficients to assess both test-retest reliability (in 16.67% of studies, or seven studies) and inter-rater reliability (in 9.52% of studies, or four studies). Pearson's correlation coefficient was used for parallel form reliability (4.76%) in two studies, while 19.05% did not specify the reliability method. This diversity in psychometric approaches underscores the multidimensional nature of nursing research, emphasizing the importance of standardizing reporting practices to enhance transparency and replicability. Advanced statistical

Table 1. Overview of Study Corpus: Journal Titles, Years of Publication, and Number of Selected Articles

Journal	Year (No.)		Total
	2021	2022	
Clinical simulation in nursing	1	5	6
European journal of oncology nursing	2	3	5
International journal of nursing studies	3	3	6
Journal of cardiovascular nursing	3	2	5
Journal of nursing scholarship	2	3	5
Journal of professional nursing	2	3	5
Nurse education today	2	4	6
Nursing outlook	2	2	4

Table 2. Overview of Psychometric Properties in Nursing Research Articles (N = 42)

Psychometric Property and Type	Method/Statistical Technique	No. (%)
Validity		
Construct validity	Factor analysis	19 (45.24)
Internal structure	Confirmatory factor analysis, exploratory factor analysis	12 (28.57)
Content validity	iterative expert assessments, literature searches, and pilot studies	8 (19.05)
Face validity	expert panels and pilot tests	7 (16.67)
Convergent validity	Confirmatory factor analysis, discriminant validity testing	6 (14.29)
Psychometric evaluation	Confirmatory factor analysis, discriminant validity testing	5 (11.90)
Concurrent validity	Pearson's correlation coefficient	3 (7.14)
Discriminant validity	Confirmatory factor analysis, discriminant validity testing	3 (7.14)
Criterion validity	Correlation coefficients	3 (7.14)
Predictive validity	Pearson's correlation coefficient	2 (4.76)
Response processes	Defined in the method section	2 (4.76)
Validation processes (including translation and validation)	Defined in the method section	2 (4.76)
Literature-based validation	Defined in the method section	2 (4.76)
Consequential validity	Structural equation modeling	1 (2.38)
No Specific validity mentioned		16 (38.10)
Reliability		
Internal consistency	Cronbach's alpha	21 (50.00)
Test-retest	Test-retest correlation coefficient	7 (16.67)
Inter-rater	Inter-rater correlation coefficient	4 (9.52)
Parallel form reliability	Pearson's correlation coefficient	2 (4.76)
No specific reliability mentioned		8 (19.05)

techniques, such as Structural Equation Modeling, highlight a growing sophistication in psychometric evaluation (Table 2). Encouraging researchers to align their psychometric choices with study requirements ensures a robust foundation for interpreting results in nursing research.

The studies examined exhibit a notable trend in reporting either validity or reliability measures, with distinct emphases on the psychometric properties of the instruments employed. In the subset of studies that exclusively reported validity without explicit mention of

reliability, various investigations focused on establishing the credibility of their instruments. For instance, one provided information on the validity of specific subscales without referencing reliability, and another reported the test as valid and reliable without detailing reliability measures for the study. These studies collectively underscore a concentration on the validation of instruments, utilizing methods such as construct validity and internal reliability assessment.

Conversely, a distinct set of studies concentrated solely on reliability, with minimal mention of validity

Table 3. Distribution of Psychometric Property Combinations in Studies (N = 42)

Type of Psychometric Property	No. (%)
Content validity, construct validity (factor analysis), internal consistency (Cronbach's alpha)	15 (35.71)
Face validity, content validity, construct validity (confirmatory factor analysis), internal consistency (Cronbach's alpha)	6 (14.29)
Face validity, content validity, internal consistency (Cronbach's alpha), construct validity (factor analysis)	10 (23.81)
Internal consistency (Cronbach's alpha) and reliability	11 (26.19)
Construct validity (factor analysis) and reliability (Cronbach's alpha)	6 (14.29)
Face validity, content validity, and construct validity, with a focus on reliability (intraclass correlation coefficient)	2 (4.76)
Face validity, convergent validity, divergent validity, internal reliability (Cronbach's alpha)	4 (9.52)
Face validity, internal consistency (Cronbach's alpha), and validity of the scales	1 (2.38)

measures. Noteworthy examples include a research, which provided alpha values for certain dimensions without explicitly mentioning validity measures, and another one, which emphasized high internal reliability but did not specify validation procedures. This subset emphasizes the importance of establishing the consistency and dependability of measurements without delving extensively into validation techniques. The identified studies collectively highlight the varying methodological priorities within the research landscape, showcasing a nuanced approach to psychometric property reporting in the reviewed literature.

From the provided data, some studies reported both validity and reliability. By analyzing the keywords used in the studies, we can classify them into groups based on the research methods and statistical techniques they employed. Here's a classification in [Table 3](#):

The presented results provide a detailed breakdown of psychometric property combinations in nursing RAs, expressed in percentages for better clarity. The most prevalent combination remains Content Validity, Construct Validity through Factor Analysis, and Internal Consistency measured by Cronbach's alpha, accounting for 35.71% of the studies. The second most common profile, consisting of Face Validity, Content Validity, Construct Validity through Confirmatory Factor Analysis, and Internal Consistency measured by Cronbach's alpha, is observed in 14.29% of the studies, showcasing methodological diversity with a focus on validation techniques.

Notably, 23.81% of studies adopt an approach emphasizing Face Validity, Content Validity, Internal Consistency (Cronbach's alpha), and Construct Validity through Factor Analysis, demonstrating a comprehensive validation strategy. The data further

highlight the significance of Internal Consistency and Reliability (Cronbach's alpha), which are jointly considered in 26.19% of instances. The findings also indicate a subset of studies (14.29%) focusing on both Construct Validity through Factor Analysis and Reliability, while a smaller proportion of 4.76% concentrates on Face Validity, Content Validity, Construct Validity, with a specific focus on Reliability measured by the Intraclass Correlation Coefficient.

Finally, a singular study (2.38%) explores Face Validity, Internal Consistency (Cronbach's alpha), and the overall Validity of the scales, underscoring the diversity in methodological approaches within the reviewed literature. Overall, the results suggest a nuanced and multifaceted consideration of psychometric properties, reflecting a commitment to robust research practices across diverse fields.

4. Discussion

The genesis of this study was motivated by the pronounced emphasis on the validation and reliability evidence of instruments in Nursing RAs. The revealed results illuminate a disconcerting trend, indicating a substantial deficit in the reporting of these crucial elements. Some of the scrutinized papers neglected to report reliability and validity, with 38.10% failing to report validity, and 19.05% omitting reliability measures. This absence of reporting echoes the concerns raised by recent studies (20), highlighting the potential impact on the ability of authors and readers to intelligently gauge the extent to which measurement errors affect results and interpretations.

The comprehensive overview of psychometric properties in nursing RAs further substantiates the concerns raised regarding the inadequacy of reporting. The dominance of construct validity, explored in 45.24%

of studies through Factor Analysis, reflects a commitment to understanding the underlying structures of measured constructs. However, the revealed deficits in reporting, as indicated by Tazik (20), suggest a prioritization of journal publication conventions over theoretical prerequisites. The lack of detailed explanations for translation processes and item changes raises concerns about the validity and reliability of the employed instruments.

The implications of inadequate reporting are manifold. Firstly, the shortfall in reliability and validity reports within Nursing RAs suggests a prioritization of journal publication conventions over theoretical prerequisites, posing a risk of measuring incorrect constructs and compromising the integrity of conclusions. The examination of construct validity in nursing research reveals notable shortcomings, with researchers predominantly reporting various types of validity instead of placing emphasis on construct validity. This discrepancy contributes to a gap between theoretical expectations and empirical practices.

The multidimensional nature of nursing research has led to the fragmentation of instrument use and validity and reliability reporting across diverse study areas. Common trends, such as the measurement of content validity through pilot studies and internal consistency through Cronbach's alpha, highlight the challenges of ensuring comprehensive validity and reliability assessments in the absence of clear reporting trends. Notably, the tendency to prioritize reliability over validity, particularly evident in studies assessing internal consistency through Cronbach's alpha, contradicts the emphasis on both properties for ensuring the quality of nursing research findings.

Moreover, a noteworthy concern arises from the assumption that instruments validated for specific variables in particular nursing contexts can be extrapolated to any related target or context. This assumption lacks a validity check for new contexts and targets, emphasizing the importance of context specificity as advocated by recent works.

The findings collectively underscore a critical need for enhancing the rigor of validity and reliability reporting in Nursing RAs. Neglecting these essential elements jeopardizes the credibility and interpretability of nursing research findings. The observed gap between validity theory and practice in nursing research journals suggests that the reporting of specific validity evidence

is driven more by journal conventions than the purpose of measurement. This discordance may contribute to shaping theoretical conceptions of validity based on conventional practices rather than theoretical requirements.

In conclusion, researchers in the nursing field are encouraged to adopt a more transparent and comprehensive reporting approach, adhering to validity guidelines and standards. Collaboration between theorists and practitioners in nursing research is pivotal in aligning validity practices with theoretical expectations. Nursing journal editors play a crucial role in promoting reporting guidelines and could consider incorporating validity courses into graduate and post-graduate nursing curricula. Only through concerted efforts can the nursing research community elevate the quality and generalizability of its research findings, fostering a more robust and theoretically sound foundation for the field. The identified weaknesses in reporting should serve as a catalyst for these improvements, ensuring that future nursing research maintains the highest standards of validity and reliability reporting.

4.1. Limitations and Future Directions in Nursing Research

Despite the comprehensive classification of psychometric properties in nursing RAs, certain limitations must be acknowledged. One notable limitation is the reliance on reported information within RAs. Although we aimed for a comprehensive analysis using the available information, the fact that researchers report their own methods in published articles can lead to potential biases or missing information. To address this limitation, future research in the nursing context may benefit from employing a combination of content analysis and direct communication with authors. Direct communication with authors would allow us to explore the research methods in more detail, resulting in a richer understanding of how psychometrics are actually used in nursing studies.

Moreover, it is essential to recognize that the focus of this study was on a specific timeframe and a selective number of journals. This deliberate scope may limit the generalizability of the findings to the broader landscape of nursing research. For future research to be more generalizable (have higher external validity), it's important to consider including more recent studies

and publications from a wider variety of nursing journals. The evolving nature of psychometrics in the nursing field necessitates an ongoing exploration of contemporary studies to capture the latest trends and advancements in reliability and validity assessments.

4.2. Conclusions

In conclusion, this study highlights the critical importance of methodological rigor in assessing reliability and validity within nursing research. By offering a contemporary perspective, the synthesis of findings provides valuable insights into prevailing trends while identifying specific areas for refinement. As the nursing field continues to evolve, researchers are encouraged to integrate recent advancements in psychometrics and embrace a holistic approach to reliability and validity assessments.

The study underscores the importance of refining research instruments to withstand the scrutiny of scientific evaluation, contributing meaningfully to the advancement of knowledge in diverse nursing domains. The call for advancing methodological standards in contemporary nursing research is imperative, and researchers should actively engage with the latest developments in psychometrics to ensure the robustness and reliability of their research instruments. Through these concerted efforts, the nursing research community can foster a culture of methodological excellence, ultimately enhancing the quality and impact of research outcomes in the field.

Acknowledgements

This work was under the support of Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran (U-98281), with an approved ethics code (IR.AJUMS.REC.1398.956).

Supplementary Material

Supplementary material(s) is available [here](#) [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

Authors' Contribution: Conceptualization: K. T.; data curation: K. T., and M. M.; formal analysis: K. T., and M. M.;

funding acquisition: K. T., and M. M.; investigation: K. T., and M. M.; methodology: K. T., and M. M.; project administration: K. T., and M. M.; resources: K. T., and M. M.; software: K. T.; supervision: K. T.; validation: K. T., and M. M.; visualization: K. T., and M. M.; writing – original draft: K. T., and M. M.; writing – review and editing: K. T., and M. M.

Conflict of Interests Statement: The authors declare no conflicts of interest.

Ethical Approval: IR.AJUMS.REC.1398.956

Funding/Support: Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran (U-98281).

References

1. Dörnyei Z. *Research methods in applied linguistics*. Oxford, UK: Oxford University Press; 2007.
2. Chan EKH. Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments. In: Zumbo BD, Chan EKH, editors. *Validity and Validation in Social, Behavioral, and Health Sciences*. Cham: Springer International Publishing; 2014. p. 9-24. https://doi.org/10.1007/978-3-319-07794-9_2.
3. Stachl C, Pargent F, Hilbert S, Harari GM, Schoedel R, Vaid S, et al. Personality Research and Assessment in the Era of Machine Learning. *Europ J Personal*. 2020;**34**(5):613-31. <https://doi.org/10.1002/per.2257>.
4. Klein JT. Evaluation of interdisciplinary and transdisciplinary research: a literature review. *Am J Prev Med*. 2008;**35**(2 Suppl):S116-23. [PubMed ID: 18619391]. <https://doi.org/10.1016/j.amepre.2008.05.010>.
5. Gethmann CF, Carrier M, Hanekamp G, Kaiser M, Kamp G, Lingner S, et al. *Interdisciplinary Research and Trans-disciplinary Validity Claims*. Cham: Springer; 2015. <https://doi.org/10.1007/978-3-319-11400-2>.
6. Rieh SY, Danielson DR. Credibility: A multidisciplinary framework. *Ann Rev Inf Sci Tech J*. 2008;**41**(1):307-64. <https://doi.org/10.1002/aris.2007.1440410114>.
7. Cruz Rivera S, Aiyegbusi OL, Ives J, Draper H, Mercieca-Bebber R, Ellis C, et al. Ethical Considerations for the Inclusion of Patient-Reported Outcomes in Clinical Research: The PRO Ethics Guidelines. *JAMA*. 2022;**327**(19):1910-9. [PubMed ID: 35579638]. <https://doi.org/10.1001/jama.2022.6421>.
8. Bachman LF. *Fundamental Considerations in Language Testing*. 7th ed. Oxford, UK: Oxford University Press; 1990.
9. Williams JM. *Writing Quality Teacher-Made Tests: A Handbook for Teachers*. Livingston, New Jersey: ERIC Document Reproduction Service; 1991.
10. Elliott R, Fischer CT, Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol*. 1999;**38**(3):215-29. [PubMed ID: 10532145]. <https://doi.org/10.1348/014466599162782>.
11. Onwuegbuzie AJ, Leech NL. Taking the "Q" Out of Research: Teaching Research Methodology Courses Without the Divide Between Quantitative and Qualitative Paradigms. *Qual Quant J*. 2005;**39**(3):267-95. <https://doi.org/10.1007/s1135-004-1670-0>.

12. Bachman LF. *Statistical Analyses for Language Assessment*. Cambridge, UK: Cambridge University Press; 2004. <https://doi.org/10.1017/cbo9780511667350>.
13. Weir CJ. *Language Testing and Validation*. Hampshire, UK: palgrave macmillan; 2005. <https://doi.org/10.1057/9780230514577>.
14. Eisinga R, Grotenhuis MT, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *Int J Public Health*. 2013;**58**(4):637-42. [PubMed ID: 23089674]. <https://doi.org/10.1007/s00038-012-0416-3>.
15. Vacha-haase T, Ness C, Nilsson J, Reetz D. Practices Regarding Reporting of Reliability Coefficients: A Review of Three Journals. *J Experimet Edu*. 1999;**67**(4):335-41. <https://doi.org/10.1080/00220979909598487>.
16. Messick S. Validity. In: Linn RL, editor. *Educational measurement*. 3rd ed. New York: American Council on Education and Macmillan; 1989.
17. Chinni ML, Hubley AM. A Research Synthesis of Validation Practices Used to Evaluate the Satisfaction with Life Scale (SWLS). In: Zumbo BD, Chan EKH, editors. *Validity and Validation in Social, Behavioral, and Health Sciences*. Cham: Springer International Publishing; 2014. p. 35-66. https://doi.org/10.1007/978-3-319-07794-9_4.
18. Zumbo B. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Vancouver, Canada: University Brit Columbia; 1998.
19. Zumbo BD, Chan EKH, Chen MY, Zhang W, Darmawanti I, Mulyana OP. Reporting of Measurement Validity in Articles Published in Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement. In: Zumbo BD, Chan EKH, editors. *Validity and Validation in Social, Behavioral, and Health Sciences*. New York, US: Springer International Publishing; 2014. p. 27-34. https://doi.org/10.1007/978-3-319-07794-9_3.
20. Tazik K. Validity and Reliability Reports in Applied Linguistics Research Articles: The case of tests and questionnaire. *Int J Foreign Lang Teach Res*. 2019;**7**(28).