DOI 10.51422/ren.v22i3.439

Instituto Nacional de Neurología y Neurocirugía Manuel Velasco Suárez | Revista de Enfermería Neurológica

Test

Validación y adaptación de instrumentos psicométricos para el avance de la investigación en enfermería

Validation and adaptation of psychometric instruments for the advancement of nursing research

Erick Landeros Olvera, Ana Laura Morales Cruz, Erika Lozada Perezmitre, Rosa María Galicia Aguilar, Guadalupe Antonio González

Resumen

Introducción: de acuerdo con los análisis bibliométricos realizados a investigaciones publicadas por profesionales de enfermería, un gran porcentaje utilizan instrumentos sin validación, lo cual indica que los trabajos no son confiables, existe la posibilidad de presentar sesgos o errores. Existen diversas metodologías para crear o validar instrumentos previos, pero se requiere de una propuesta para la ciencia del cuidado, ya sea en una adaptación o creación de algún indicador empírico.

Objetivo: presentar una propuesta con la finalidad de validar y adaptar instrumentos psicométricos para el desarrollo de la investigación en enfermería.

Desarrollo: la validación de un instrumento que depende del tiempo, objetivo de estudio y de las etapas a cubrir; por lo cual, se proponen ocho etapas. 1. Traducción del instrumento original al idioma español. 2. Adaptación semántica. 3. Validación por jueces. 4. Corrección de estilo. 5. Prueba piloto. 6. Propiedades psicométricas. 7. Prueba final o análisis factorial, y 8. Prueba de sensibilidad.

Conclusión: existen más formas de validar los instrumentos, la expuesta en este trabajo, representa una forma pragmática para adaptar un instrumento al contexto donde enfermería realiza investigación; en consecuencia, obtener instrumentos que den confiabilidad a los resultados de las intervenciones de enfermería.

Palabras clave: estudio de validación, evaluación de instrumentos de investigación, enfermería.

Citation: Landeros Olvera E, Morales Cruz AL, Lozada Perezmitre E, Galicia Aguilar RM, Antonio González G. Validation and adaptation of psychometric instruments for the development of nursing research. Rev Enferm Neurol. 2023;22(3): pp. 281-290.

Correspondence: Erick Landeros Olvera
Email: erick.landerosolvera@viep.com.mx
Benemérita Universidad Autónoma de Puebla, Facultad de Enfermería

Received: August 3, 2023
Accepted: November 11, 2023

Abstract

Introduction: According to bibliometric analyses of research published by nursing professionals, a large percentage use instruments without validation, which indicates that the studies are unreliable and may present biases or errors. There are various methodologies to create or validate existing instruments; however, a proposal is needed for the science of care, whether through the adaptation or creation of an empirical indicator.

Objective: To present a proposal aimed at validating and adapting psychometric instruments to support the development of nursing research.

Methodology: tthe validation of an instrument depends on timeframe, study's objective and the stages to be covered; for which eight stages are proposed. 1. Translation of the original instrument into Spanish language. 2. Semantic adaptation. 3. Validation by judges. 4. Style correction. 5. Pilot test. 6. Psychometric properties. 7. Final test or factorial analysis, and 8. Sensitivity test.

Conclusion: There are more ways to validate instruments; the one presented in this paper represents a pragmatic way to adapt an instrument to the context where nursing conducts research; consequently, obtaining instruments that provide reliability to the results of nursing interventions.

Keywords (MeSH): validation study, research instrument evaluation, nursing.

Introduction

According to bibliometric analyses of Mexican nursing journals, over the period of 2005-2015, the average number of psychometric instruments that did not report reliability or validity was 20%.[1,2] In many cases, these are Anglo-Saxon instruments translated by the same researchers or instruments in Spanish; however, without previous validation or application in the Mexican context. In other cases the researchers constructed them to measure the variable of interest. In either case, they represent sources of internal invalidation.[3] That is, 20% of the research papers published in Mexican nursing journals from 2005 to 2015, are unreliable due to the possibility of presenting biases or type III errors.[4]

There is no single technique or methodology for adapting or validating an instrument. It all depends on the objective of the study. However a long and rigorous methodological process is required, and its design is usually referred as multistage or process-oriented.[5] Literature provides a great deal of information based on three types of validation: content, construct, and criterion.[6]

For content validation, qualitative, quantitative, and mixed evaluations are included. The first one includes the Delphi method and the Fehring model[7]. In the second one, peer review is measured, considering the content validity ratio, the Kendall's "*W*" concordance coefficient, content validity index, and Aiken's V test.[8] The third one uses the Q methodology that encompasses both research paradigms.[9]

As for construct validity, it can be evaluated

through an exploratory or confirmatory factor analysis, in addition to the use of the VARIMAX method.[10]

Finally, the criterion validity, which includes statistical approaches such as the Kappa concordance coefficient, sensitivity, ROC curve, parametric correlations, and interclass correlation coefficient, will depend on the type of variable.

According to the above, for nursing science, *"the method"* should not be different, however, there is no reference for this discipline. There is no single method that is better than another, it depends on pragmatic considerations. Therefore, nursing professionals need to have valid and reliable empirical indicators. In this way, the potential for errors is reduced.[11] Given the above arguments, the objective of this manuscript is to present a proposal to validate and adapt psychometric instruments for the development of nursing research.

Validation and adaptation of instruments

The design for the validation of an instrument is of longitudinal process and scope.[5] The following eight stages are proposed:

1. **Traduction:** If the instrument is not written Spanish, the items should be translated into Spanish. The documents that constitute the instrument must be translated by a certified translator in the language in which the instrument is originally written. In-house translations, the use of free translators on the internet, and the use of artificial intelligence (AI) should be avoided.[12], The first version of the instrument is obtained should follow these criteria.

Sometimes it is necessary to request permission from the original authors to make the corresponding adaptation or to request psychometric details of the instrument. It is always a matter of professional courtesy to communicate directly with the original authors and to ensure that the instrument is truly public or to obtain the consent of the copyright holder.[13]

2. **Semantic adaptation of the translation.** At this stage, it is important to make decisions about the meaning of each item; but first, whether they are questions or statements must be verified, in order to link them to the response format. For example: if the sentences need to be in the affirmative sense, the verb is placed in the present indicative, to be congruent with the Likert-type response format.[14]

It is recommended that the meaning of the original sentence of each item be preserved, despite the changes in syntax; consequently, the second version of the instrument should be achieved.

3. **Face validity through peer-review technique.** Ten experts in the subject related to the instrument are invited, with a consideration of 50% of attrition since almost always half of the specialized peers do not submit their evaluations or postpone it.[15]

The expertise of the peers should be discussed with the purpose of unifying competencies, experience in the subject, and academic qualifications, among others. The invitation to each judge is made by an official letter through institutional mail, where they will be given an unpublished instrument, and the provided information must be kept confidential. If the judge accepts, it is explained to they that their evaluation has two phases. The quantitative phase, which consists of providing them with an evaluation format using a Likert-type scale, evaluating each item individually, ranging from 0 = definitely not related, 1 = Not related, 2 = Not sure of their relationship, the items require further review, 3 = Related, but minor modifications are necessary, and 4 = Extremely related. In order to evaluate each item according to the previous

scale, the nominal definition of the notion to be evaluated must be placed in the instructions for the judges, allowing them to observe the degree of correlation. For example: the vocation of service for human care is defined as the *'inclination or sense of inspiration that the student possesses to offer or dedicate nursing care to healthy or ill individuals'*.[16] With the definition, an evaluation form is sent with all the items, see the example in Table 1, where some of the items that make up the instrument of vocation of service to human care in nursing students are presented.[16]

Table 1. Example of judges' evaluation form

Ítems	*0*	*1*	*2*	*3*	*4*
7. I promote health with my family or friends.					
9. I like to listen attentively to people when they have health problems.					
12. I am happy to offer my nursing services in a community with low social development.					

Source: Validation of the human care service vocation instrument in nursing students[16]

Accordingly, the ten evaluation forms from the judges are entered in a data table, those items with a score equal to or greater than three are considered acceptable (A), whereas those with a score lower than three are classified as not acceptable (NA). An example is presented in Table 2 (only a portion of a 23-items instrument of is included).

Table 2. Judges´ evaluation data

	JUDGES										*NA*	*A*
	1	2	3	4	5	6	7	8	9	10		
ÍTEMS												
1	3	4	4	4	4	3	4	4	4	4 ⇨	0	10
2	3	3	3	2	4	3	3	2	2	4 ⇨	3	7
…23	4	4	4	3	2	3	3	4	4	4 ⇨	1	9
Total											↓ 48	↓ 182

Own source. Note: The criterion for identifying the NA (Not Acceptable) and A (Acceptable) items is determined horizontally, considering each of the judges (⇨), and then summed vertically. This takes into account all the items (↓).

Employing the results from 10 judges, the following parameters are calculated: item validity index (IVI), criterion validity (CV), and content validity index (CVI).[17]

Considering the results of the previous table as an example, the IVI is calculated with the sum of the items with a score lower than three points (NA= 48) and divided by the total number of judges (10). This is done in the following way: (48/10) IVI= 4.8. This result is divided by the total number of items (23), to obtain the validity criterion: (4.8/23) CV= .2086. The result is

multiplied by 100 and represents a 20% possibility of error that the items do not measure the concept they claim to measure. To obtain the CVI, the items with scores equal to or greater than three were added (A= 182) and divided by the total number of items (182/23), resulting in 7.91, which is close to 10. The results in the three indexes of this exercise are acceptable to validate analysis from judges in its quantitative phase.

In the second phase, the judges make qualitative observations of the instrument, and the suggestions of each judge must be discussed to decide the pertinence of the modifications to the item. Each change must be supported not only by the judge's opinion but also by the published evidence for the adaptation of the item in the context where the instrument will be applied or tested. In this way, the third version of the instrument is obtained.

4. **Style correction.** The result of the previous stage must be reviewed by a proofreader (expert in Spanish language linguistics) to improve the syntax and coherence of the sentences of each item, thus obtaining the fourth version of the instrument.

5. **Pilot test.** To test the fourth version of the instrument, a Pilot Test (PT) should be applied to a sample that meets the inclusion criteria to measure the concept of interest according to the objective of the study.[18] The characteristics should be similar; for example, sociodemographic variables (gender, age, schooling, marital status, students, professionals), anthropometric variables (weight, BMI, percentage of fat, waist circumference), and clinical variables (blood pressure, glucose, lipids, clinical condition). The objective of a PT is to know the feasibility of the instrument, the response time, and the clarity of the wording of the items.[19,20] Some authors recommend performing a PT on a minimum of 30 study subjects to test an instrument, the reason being that this sample size, although small, has a sufficient statistical effect to determine whether or not there is a normal distribution in the data curve and preliminary reliability of the instrument.[21] In the PT, study subjects are allowed to make observations on any item, these observations can be considered for developing changes and obtain the fifth version of the instrument.

6. **Psychometric properties.** In the fifth version of the instrument, the preliminary psychometric properties are established. Psychometrics is the science in charge of measuring cognitive concepts -usually subjective-, which require precise measurements of the constructs of a concept.[5] The psychometric properties suggested to be established are the following: number of items (items, statements, or questions), maximum and minimum values according to the response score, and a five-choice Likert-type format (specify the response options).[22] Similarly, the dimensions or sections of the instrument (if applicable) should be established and the number of items that make up each section should be specified, as well as the cut-off points to classify the grade to be measured. Here the preliminary internal consistency of the PT is placed (Cronbach's alpha). [21]

7. **Final testing.** This stage consists of testing the fifth version of the instrument on a significant sample, robust enough to obtain an acceptable effect size and valid and reliable results. To achieve the sixth and final version of the instrument, the statistical test pursued at this stage is factor analysis (FA), a multivariate technique intended to identify the structure between a group of variables to be analyzed (items) that underlie a group of data or variables, which in the instrument may be represented by sections or dimensions (hereafter referred to as factors).[10]

With factor analysis, it is possible to explore and confirm whether by rotating the items of an instrument, they are grouped - from a statistical point of view - in the dimension where the researcher originally grouped the items - from a theoretical point of view.[23]

There are two main techniques: exploratory FA and confirmatory FA. The first allows the identification of items that are more closely interconnected with one factor than with another. The second, after having explored and determined the set of items that belong to each factor, as its name indicates, confirms an already determined or pre-established number of items that make up each factor.[24]

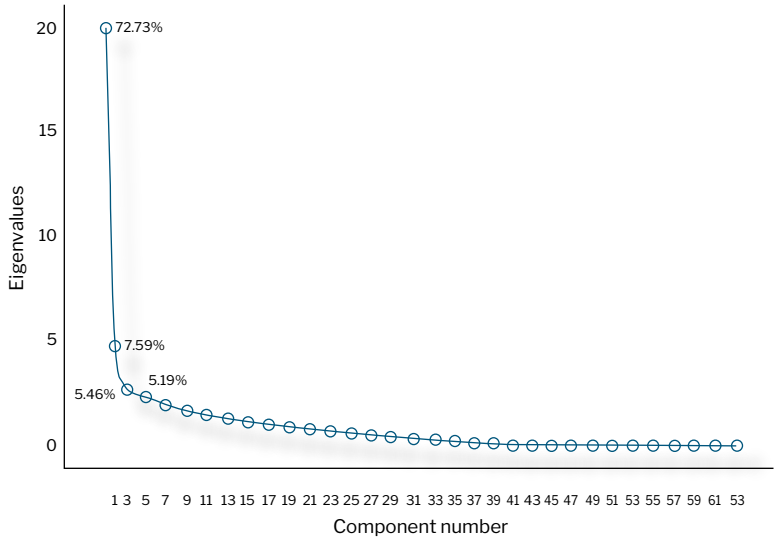
For exploratory FA, a correlation matrix is always used to identify the relationship between each data item and each factor (it must have an effect size greater than .30), and for confirmatory FA, structural equation models are used.[10]

For both techniques, the postulates or requirements for applying the technique must be taken into account: normality of the data, homoscedasticity of the variance, homogeneous samples (subjects with equivalent characteristics), without collinearity; in addition to the prior application of the Kaiser-Meyer-Olkin sample test (with values close to unity) and Bartlett's test of sphericity (with a $p<.05$), which prove the absence of significant correlations between the variables.

It should be considered that, in samples smaller than 150 study subjects, it does not support either of the two factor analyses.25 The sample should be a minimum of 300 instruments; some authors recommend 20 to 30 study subjects for each item to be analyzed, in other words, if the instrument has 20 items, a sample of 400 study subjects answering the instrument is required to submit it to a FA.[18]

In factorial analyses, it is necessary to generate an orthogonal rotation matrix. Usually, the varimax rotation is performed, which allows the data to be rotated to observe the variance explained through a sedimentation graph, thus clarifying the percentage of variance in each item in a cascade. For example, in graph I, it can be identified that the first four items explain more than 70% of the variance of the instrument, the others are almost in a horizontal line, which denotes consistency among the items.

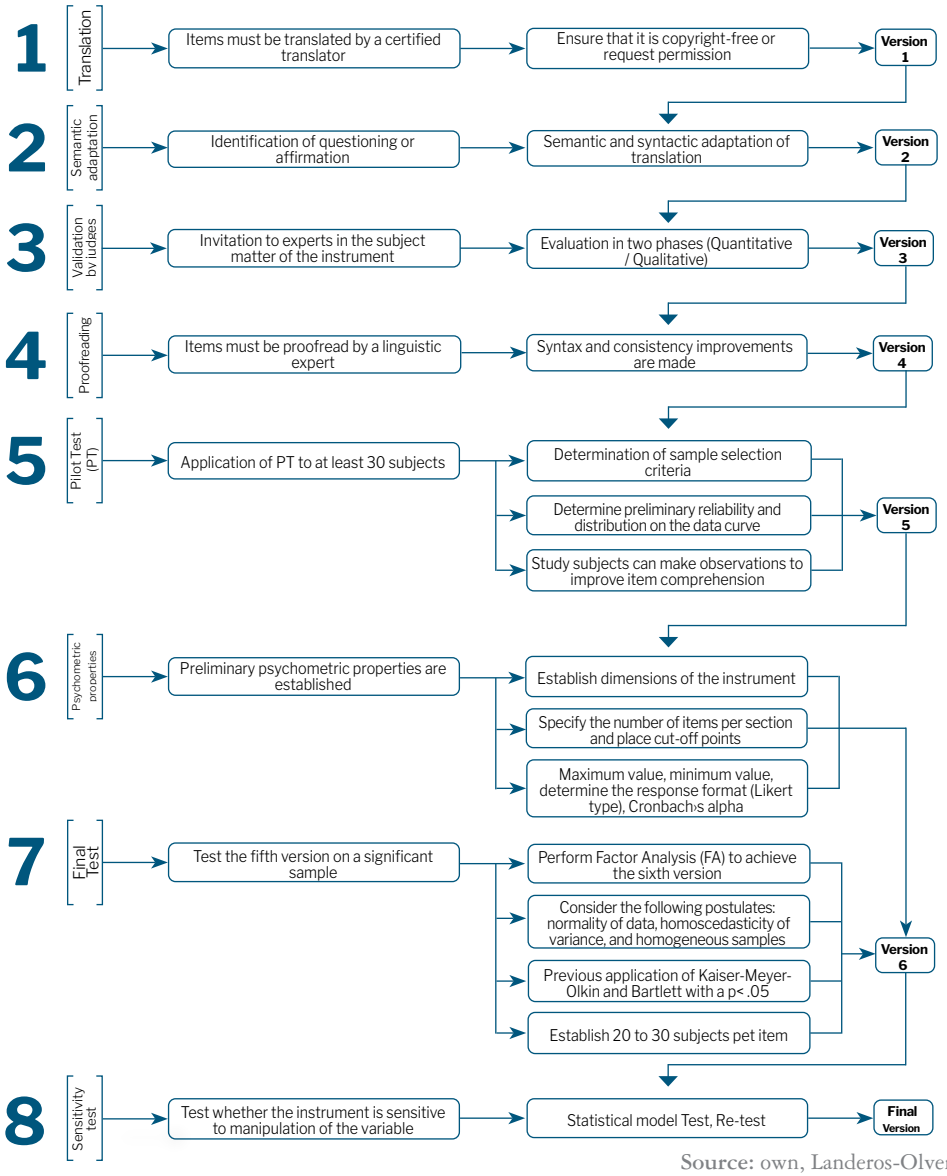
Figure 1. Sedimentation graph



Source: Validation and standardization of the instrument: evaluation of caregiving behaviors in Mexican nurses.[26]

8. **Sensitivity test.** The authors of this paper call it this way because it proves that the instrument is sensitive to the manipulation of the variable and, therefore, to time. With an experimental design under a statistical model *test, retest*,[5] where the variable to manipulate is the concept that the instrument claims to measure, after the intervention (or intermediate measurements), in the end, the scores of the variable should change. To achieve this, it is necessary to comply with the characteristics of a clinical trial,[26] to ensure greater certainty in the control of confounding variables to ensure that the changes in the variable are due to the manipulation through the nursing intervention. If after the manipulation, in the retest, the difference is statistically significant ($p < .05$), congratulations, your instrument is valid, reliable, and sensitive to the manipulation of the concept it claims to measure.

Figure 2. Process of validation and adaptation of psychometric instruments



Source: own, Landeros-Olvera et al. (2023)

Procedure and final recommendations

Many times, when there are no instruments to measure a particular concept, the option is to create one, but this requires several steps, such as a thorough review of the literature, perhaps previously analyzing the concept that the instrument measures, and a sharp handling of theoretical connotations around the concept to be measured. Then, to create the theoretical statements or items with their corresponding dimensions, after having this first version, the instrument would have to be submitted from stage 2 which has been shown in this paper.

Within the psychometric properties, a 5-point Likert-type scale is recommended, to avoid the *"ceiling-floor"* phenomenon[27] or to invert the sense of the score in some items to ensure the attention of the subject answering and avoid susceptible error.[11]

In such a case, it is recommended that the cut-off points be made based on indices from zero to 100, using quartiles (0-25, 26-50, 51-75, and 76-100). This is equivalent to Bad, Fair, Good, and Excellent, thus speeding up statistical interpretation.

If stage eight is carried out to test the sensitivity of the instrument to the manipulation of the variable, the nursing intervention should be constructed based on the premises of a prior systematized theory and review of the literature.[28]

It is not suggested to validate instruments with dichotomous response scales; however, if necessary, it is suggested to consult the Kurder-Richardson statistical validation (KR-20 or KR-21) for this type of scale.[29]

If the instrument exists, but has not been tested in your country, validate it; if it does not exist, construct and validate it, avoid using unreliable instruments in research or graduate theses.

Conclusion

This academic manuscript presents a proposal to validate and adapt psychometric instruments for the development of nursing research; however, it is not the only way it can be done. The methodology depends on the type of instrument, the objectives of the study, the subject matter to be addressed, the resources available, the clarity of the concept to be measured and its constructs, and whether the validation corresponds to an existing instrument in another language or another Hispanic country, or whether it is the original construction of a new instrument.

Finally, your thesis or research project may be the adaptation and validation of some psychometric instrument that helps to measure nursing care, this would imply a great contribution.

References

1. **García RM, Gómez ÁMG, Aguilar PI, Pérez MGP, Velázquez DL, Soriano SM, et al.** Trends and features of published nursing research in Mexico. Enferm Univ [Internet]. 2011 [Accessed 15 Jul 2023];8(1). Available at: https://cutt.ly/dwMu5Zsc
2. **Landeros-Olvera E, Ramírez-Girón N, Yáñez-Lozano Á, Guzmán-Ramírez G, Galicia-Aguilar RM.** Topics and methodological characteristics of nursing research published in Mexican journals: 2010-2015. Enferm Univ [Internet]. 2018 [Accessed 15 Jul 2023];15(3). DOI http://dx.doi.org/10.22201/eneo.23958421e.2018.3.65994
3. **Hernández-Sampieri R, Mendoza C.**

Research methodology. Quantitative, qualitative and mixed routes. Mexico City: Mc Graw Hill Education; 2018.

4. **Villasís-Keever MÁ, Márquez-González H, Zurita-Cruz JN, Miranda-Novales G, Escamilla-Núñez A.** Research protocol VII. Validity and reliability of the measurements. Rev Alerg México [Internet]. 2018 [Accessed 15 Jul 2023];65(4). Disponible en: https://doi.org/10.29262/ram.v65i4.560
5. **Medrano L, Pérez E.** Manual de psicometría y evaluación psicológica. Argentina: Brujas; 2019.
6. **Cruz-Avelar A, Cruz-Peralta ES.** Methodology for the construction of measurement instruments in health. Alerg Asma Inmunol Pediátricas [Internet]. 2018 [Accessed 15 Jul 2023];26(3). Available at: https://cutt.ly/fwMiwmFy
7. **López FR, Avello MR, Palmero UDE, Sánchez GS, Quintana ÁM.** Validation of instruments as a guarantee of credibility in scientific research. Rev Cuba Med Mil [Internet]. 2019 [Accessed 15 Jul 2023];48. Disponible en: https://cutt.ly/OwMietbq
8. **Sarabia CCM, Alconero CAR.** Keys for the design and validation of questionnaires in health sciences. Enferm Cardiol [Internet]. 2019 [Accessed 15 Jul 2023];26(77). Available from: https://cutt.ly/mwMieAOA
9. **Urrutia EM, Barrios AS, Gutiérrez NM, Mayorga CM.** Optimal methods for content validity. Educ Médica Super [Internet]. 2014 [Accessed 16 Jul 2023];28(3). Available at: https://cutt.ly/dwMie1yG
10. **Ávila MM.** Confirmatory factor analysis: a knowledge management model in the public university. RIDE Rev Iberoam Investig Desarro Educ [Internet]. 2021 [Accessed 15 Jul 2023];12(23). DOI: https://doi.org/10.23913/ride.v12i23.1103
11. **Blázquez SN.** Validated Questionnaires: Solid Tools for Research. Actas Dermo-Sifiliográficas [Internet]. 2017 [Accessed 15 Jul 2023];108(10). Available from: https://doi.org/10.1016/j.ad.2017.09.005
12. **Corvalán JG.** Artificial intelligence: challenges and opportunities – Prometea: the first artificial intelligence of Latin America at the service of the Justice System. Rev Investig Const [Internet]. 2018 [Accessed 16 Jul 2023];5(1):295. DOI: https://doi.org/10.5380/rinc.v5i1.55334
13. **Rodríguez CP, Casañas SR, Collado PA, Maldonado AG, Vicente GM, Escuriet PR.** Cultural Adaptation of the "*Questionnaire for Assessing the Childbirth Experience (QACE)*". Enferm Glob [Internet]. 2020 [Accessed 15 Jul 2023];19(4). Available from: https://doi.org/10.6018/eglobal.402271
14. **Lira MT, Caballero E.** Cross-cultural adaptation of evaluation instruments in health: history and reflections on why, how and when. Rev Médica Clínica Las Condes [Internet]. 2020 [Accessed 17 Jul 2023];31(1). Available at: https://doi.org/10.1016/j.rmclc.2019.08.003
15. **Zurita-Cruz JN, Villasís-Keever MÁ.** Main biases in clinical research. Rev Alerg México [Internet]. 2021 [Accessed 17 Jul 2023];68(4). Available at: https://doi.org/10.29262/ram.v68i4.1003
16. **Antonio GG, Montes AJS, Ramírez-Girón N, Landeros-Olvera E.** Validation of the instrument of vocation of service to human care in nursing students. Index Enferm [Internet]. 2021 [Accessed 17 Jul 2023];30(3):254-8. Available at: https://cutt.ly/uwMivipA

17. **Waltz FC, Strickland LO, Lenz RE.** Measurement in nursing and health research. 4ta ed. New York: Springer Publishing Company. 2010.
18. **Morales-Castillo FA, Hernández-Cruz MC, Morales RMC, Landeros-Olvera EA.** Validation and standardization of the instrument: Assessment of the provided caring behaviors in Mexican nurses. Enferm Univ [Internet]. 2016 [Accessed 17 Jul 2023]; 13(1):3-11. Available at: https://doi.org/10.1016/j.reu.2015.11.005
19. **Mora EA, Carrasco AAS, Muñoz VPM, Salinas RS, Huerta SC, Noriega EP, et al.** Characteristics of the pilot test: a review of articles published in nursing. Rev Enferm Neurológica [Internet]. 2015 [Accessed 19 Jul 2023];14(3):169-75. DOI: https://cutt.ly/awMimi2M
20. **Muñiz J, Fonseca-Pedrero E.** Ten steps for the construction of a test. Psicothema [Internet]. 2019;31(1):7-16. Available at: https://doi.org/10.7334/psicothema2018.291
21. **Yunkor-Romero YK, Ochoa-Pachas JM.** Internal Validation of Research Instruments in Social Sciences. ACTA Juríd Peru [Internet]. 2020 [Accessed 19 Jul 2023];3(2):112-35. Available at: https://cutt.ly/WwMiWfKl
22. **Muñiz J.** Introducción a la psicometría teoría clásica y TRI. 1era ed. Madrid:Pirámide. 2018.
23. **Luján-Tangarife JA, Cardona-Arias JA.** Construction and validation of measurement scales in health: a review of psychometric properties. Arch Med [Internet]. 2015 [Accessed 19 Jul 2023];11(31). DOI: https://cutt.ly/2wMiW9KH
24. **Lloret-Segura S, Ferreres-Traver A, Hernández-Baeza A, Tomás-Marco I.** Exploratory Item Factor Analysis: A practical guide revised and updated. An Psicol [Internet]. 2014 [Accessed 19 Jul 2023];30(3). DOI: https://dx.doi.org/10.6018/analesps.30.3.199361
25. **Roco Videla Á, Hernández Orellana M, Silva González O.** What is the adequate sample size to validate a questionnaire? Nutr Hosp [Internet]. 2021 [Accessed 20 Jul 2023]; 38(4). DOI: https://dx.doi.org/10.20960/nh.03633
26. **Cobos-Carbó A, Augustovski F.** CONSORT 2010 Declaration: Updated guideline for reporting parallel group randomised trials. Med Clínica [Internet]. 2011 [Accessed 21 Jul 2023];137(5). Available at: https://cutt.ly/qwMiR6Rw
27. **Matas A.** Likert-Type Scale Format Design: State of Art. Rev Electrónica Investig Educ [Internet]. 2018 [Accessed 21 Jul 2023];20(1). DOI: https://doi.org/10.24320/redie.2018.20.1.1347
28. **Carvajal EY, Herrera BS.** Models of nursing applied in clinical practice: integrative review. Arch Med Col [Internet]. 2018 [Consultado 21 Jul 2023];18(1):86-96. Available at: https://www.redalyc.org/journal/2738/273856494008/html/
29. **Cascaes SF, Gonçalves E, Valdivia ABA, Bento GG, Silva CTL, Soleman HSS, et al.** Estimators of internal consistency in health research: the use of the alpha coefficient. Rev Peru Med Exp Sal Púb [Internet]. 2015 [Accessed 21 Jul 2023];32(1). Available at: https://cutt.ly/6wMiYXry