Journal of Multidisciplinary Healthcare

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Translation, Cross-Cultural Adaptation, and Validation of Measurement Instruments: A Practical Guideline for Novice Researchers

Paulo Cruchinho[1], María Dolores López-Franco[2], Manuel Luís Capelas[3], Sofia Almeida[4], Phillippa May Bennett[5–7], Marcelle Miranda da Silva[1,8], Gisela Teixeira[1], Elisabete Nunes[1], Pedro Lucas[1], Filomena Gaspar[1] On Behalf of the Handovers4SafeCare

[1]Nursing Research, Innovation and Development Center (CIDNUR) of Lisbon, Nursing School of Lisbon, Lisboa, Portugal; [2]CTS-464 Nursing and Innovation in Healthcare, University of Jaén, Jaén, Spain; [3]Universidade Católica Portuguesa, Faculty of Health Sciences and Nursing, Center for Interdisciplinary Research in Health (CIIS), Lisboa, Portugal; [4]Universidade Católica Portuguesa, Faculty of Health Sciences and Nursing, Center for Interdisciplinary Research in Health (CIIS), Porto, Portugal; [5]Center for English, Translation, and Anglo-Portuguese Studies (CETAPS), Lisboa, Portugal; [6]Faculty of Social Sciences and Humanities of the New University of Lisbon, Lisboa, Portugal; [7]Faculty of Arts and Humanities of the University of Coimbra, Department of Languages, Literatures and Cultures, Coimbra, Portugal; [8]Federal University of Rio de Janeiro, Anna Nery Nursing School, Rio de Janeiro, Brazil

Correspondence: Paulo Cruchinho, Nursing School of Lisbon, Avenida Prof. Egas Moniz, Lisboa, 1600-190, Portugal, Tel +351 217913400, Email pjcruchinho@esel.pt

Abstract: Cross-cultural validation of self-reported measurement instruments for research is a long and complex process, which involves specific risks of bias that could affect the research process and results. Furthermore, it requires researchers to have a wide range of technical knowledge about the translation, adaptation and pre-test aspects, their purposes and options, about the different psychometric properties, and the required evidence for their assessment and knowledge about the quantitative data processing and analysis using statistical software. This article aimed: 1) identify all guidelines and recommendations for translation, cross-cultural adaptation, and validation within the healthcare sciences; 2) describe the methodological approaches established in these guidelines for conducting translation, adaptation, and cross-cultural validation; and 3) provide a practical guideline featuring various methodological options for novice researchers involved in translating, adapting, and validating measurement instruments. Forty-two guidelines on translation, adaptation, or cross-cultural validation of measurement instruments were obtained from "CINAHL with Full Text" (via EBSCO) and "MEDLINE with Full Text". A content analysis was conducted to identify the similarities and differences in the methodological approaches recommended. Bases on these similarities and differences, we proposed an eight-step guideline that includes: a) forward translation; 2) synthesis of translations; 3) back translation; 4) harmonization; 5) pre-testing; 6) field testing; 7) psychometric validation, and 8) analysis of psychometric properties. It is a practical guideline because it provides extensive and comprehensive information on the methodological approaches available to researchers. This is the first methodological literature review carried out in the healthcare sciences regarding the methodological approaches recommended by existing guidelines.

Keywords: cross-cultural comparison, decision-making, psychometric properties, research design, validation studies, health services research

Introduction

Healthcare research requires the use of cross-culturally validated instruments to measure implementation of healthcare interventions and their outcomes through quantitative comparisons over time and across organizations.[1–4] The use of data obtained through culturally adapted evaluation instruments allows researchers, policymakers, managers and, health professionals to gain a more analytical view of the phenomena under study and to develop internationally accepted and recognized theories on the provision of patient care, based on the comparison of local data with broader data.[5] This approach also facilitates the identification of factors contributing to the effectiveness of healthcare intervention programs,[6] or other forms of Outcomes Research. This type of quantitative research, focused on the quality of healthcare provision, requires valid and

Received: 18 May 2023
Accepted: 21 March 2024
Published: 31 May 2024

reliable measuring instruments,[7] obtained through cross-cultural validation studies. These studies aim to confirm the capacity of measurement instruments developed in one culture to produce meaningful results when applied in another culture.[8] Measurement instruments can include questionnaires, tests, rating scales and self-reports,[9] the latter being also known as Patient-Reported Outcomes Measures (PROMs).[10]

In recent years we have conducted several cross-cultural validation studies of different measuring instruments,[11–19] which constitute a significant contribution to the development of experimental designs in the field of nursing and health services research. Several studies across different scientific areas are characterized by the use of specific terminology and by seeking to archive various equivalences across cultures. Additionally, cross-cultural validation studies involve a long and complex process that require researchers to have a wide-ranging technical knowledge of the translation, back translation, adaptation, and pre-test aspects, their purposes and options, the different psychometric properties, and the required evidence for their assessment and knowledge about quantitative data processing and analysis using statistical software. Furthermore, these studies involve specific risks of bias, which may affect the research process and results. To address these challenges, novice researchers must be well-informed about the most suitable methodological approaches.

Concepts and Specifics Terms

The adaptation and testing of measurement instruments across different international contexts over time, not only enhances their reliability and validity,[20] but also facilitates comparisons between cultures and the identification of relevant factors for developing effective interventions.[6] Cross-cultural adaptation is not limited to the translation of measurement instruments. It encompasses the adaptation and validation of these instruments in the cultural context in which they are intended to be used.[21]

Some specific terms are used in the process of cross-cultural adaptation. For example, the "target version" of a given measurement instrument is the version to be created through the process of cultural adaptation and the "target language" consists of the language into which the adaptation is intended. The "original version" is the version of the instrument that researchers intend to adapt and the "source language" is the language of the "original version". Bilingual translators in the process of cross-cultural adaptation are individuals who have a full command of both the "target language" and the "original language".[22] Translation involves converting a document from the "source language" to the "target language", considering the target audience, target culture, and the *skopos* (brief or communicative purpose).[23] In the case of translating health instruments, this encompasses factors such as accuracy, fluency, and conceptual equivalence, but also, as argued by Montalt & Davies,[24] the ethical priority of "cultural relevance", while cross-cultural adaptation comprises the identification of differences between the "source culture" and the "target culture" to maintain the equivalence of concepts. Finally, cross-cultural validation aims to ensure that the "target instrument" works as intended and has the same properties as the "original instrument".[25] Within cross-cultural validation we can distinguish the psychometric validation performed after the field testing from the validation performed during pre-testing, which aims to validate the adapted version before its exploratory use.

Types of Equivalence

The purpose of cross-cultural adaptation consists of obtaining a measurement instrument in the "target language" that is conceptually equivalent to the original. Before researchers opt for a particular methodological approach for the translation, adaptation, and cross-cultural validation of measurement instruments, it is necessary to understand the different types of equivalence that can be achieved between the "target version" and the "original version".

The equivalence can be specified in different categories varying according to the authors. Herdman et al[26] proposed a set of five categories: 1) conceptual equivalence; 2) item equivalence; 3) semantic equivalence; 4) operational equivalence and 5) equivalence of measurement. Conceptual equivalence verifies which domains and their inter-relations are important in the "target culture" for the concept of interest evaluated by the instrument. Item equivalence critically examines the items covered by the concept domains, while semantic equivalence ensures that translations of items semantically match the items in the "original version". Operational equivalence seeks to guarantee that the measurement methods used are appropriate in the "target culture" and measurement equivalence corresponds to the verification of the process result with reference to instrument's behavior related to its psychometric properties. Each one of these categories is important for judging the overall equivalence of the measurement instruments, ie their functional equivalence.[26] Peña,[27] described another equivalence categories, namely: 1) functional equivalence; 2) cultural equivalence; 3) metric equivalence and 4) linguistic equivalence. The latter corresponds to the semantic

equivalence of Herdman et al.[26] Functional equivalence assesses whether the instrument has the same behavior in both cultures. Cultural equivalence specifies how participants will answer to a given item covered by the same cultural meaning.[28] Finally, metric equivalence concerns the difficulty of a given item being expressed in two different languages.[29] According to Peña,[27] the equivalences to be obtained in the cultural adaptation of measurement instruments depend on the objectives of the studies. To establish which equivalences obtain, researchers may choose one of these two categorizations. Understanding the different categories of equivalence enables researchers to design a methodological approach for cross-cultural adaptation procedures tailored to the types of equivalence sought. If researchers adopt a standardized methodological approach proposed by an author, it also allows them to supplement the process with other procedures better suited to the characteristics of their measurement instrument and target population. This is done with the purpose of achieving or strengthening a particular type of equivalence in the instrument.

Typologies of Biases

Another element that researchers need to understand before beginning the translation, adaptation and cross-cultural validation of measurement instruments is the risk of bias. Cultural biases pose the primary threat of this process. A measurement instrument is considered biased if two or more cultural versions are inadvertently affected by an undesirable source of variance, resulting from: 1) differences in concepts between the "source culture" and the "target culture"; 2) difference between the items used to represent the constructs in the instruments and 3) the method or form of administration used.[30] Cultural biases are categorized into method bias, content bias, and construct bias based on their etiology.[31] A challenge in cross-cultural adaptation of measurement instruments is managing different response styles across cultures, namely acquiescence, ceiling and floor effects, and the tendency toward neutral responses.[32] These differences in response styles may be a source of method bias,[33] and may be more expressive in certain cultures than others and related to the need to protect the identity and privacy,[34] because of the presence of low levels of participants' motivation and the valuing of social norms of politeness.[35] Content bias can be introduced by items whose content is unfamiliar to the "target culture",[31] while construct bias occurs when there is only partial equivalence in the construct being measured between the cultures.[36]

To mitigate these cultural biases during cross-cultural adaptation, researchers can employ several strategies. One strategy is to pre-test the instrument with a sample of participants from the "target culture". Another strategy comprises conducting interviews with participants after the pre-test to assess their attributes and functioning.[30] Despite there is no robust evidence to prevent method bias, researchers may recourse to a) forced-choice response formats without middle neutral points and b) use Likert scales with an extended number of response options.[32,37,38] For instance, using 5 to 7 point response formats is deemed suitable for measuring attitudes.[39] To save time and resources, it is important that researchers identify the risk of any of these biases as early as possible, preferably before conducting pre-tests.

Methodological Approaches

The translation, adaptation and validation of instruments requires methodological guidelines developed and proposed by experienced researchers.[40] Despite this, several validation studies do not mention whether they adopted an internationally accepted guideline for their work.[41] Some authors have highlighted a lack of detailed information on the fundamentals of methodological approaches and the options available to researchers.[42] Literature reviews have also reported a lack of consensus on the methodological approaches to be followed in the process of translation, adaptation and cross-cultural validation.[25,42–44] Cha et al[6] attributed this lack of consensus not only to the specificity of research questions but also to the research environment, namely the accessibility and availability of bilingual translators. Farina et al[45] have recently shown that rigorous and pragmatic cross-cultural adaptation can be achieved with limited resources. Faced with a lack of consensus, Epstein et al[25] recommended choosing methods that best suit the context in which the evaluation instrument will be used. Furukawa et al[46] noted that this choice depends on research objectives, the availability of translators, budget, and time constraints. Additionally, Helmich et al[47] advocated that in order to produce results that truly reflect the context, the choice of methods must align with the epistemological position of the researchers.

Despite the lack of consensus, guidelines share some common elements. In a literature review carried out by Acquadro et al,[44] it was found that in order to cross-culturally adapt the PROMs, the guidelines have in common a multi-step and centralized process, at least one translation and some kind of pre-test. Regarding the questionnaires in general,

Epstein et al[25] observed that most guidelines recommend an Expert Committee, Focus Groups, and back translation of the instrument.

Guidelines should cover not only translation and cross-cultural adaptation but also psychometric validation. Some reviews have reported a lack of knowledge about the psychometric properties of adapted measurement instruments,[48–50] and incomplete information on all the psychometric validation domains.[51] For example, Danielsen et al[52] found that the psychometric properties of adapted versions validated with different tests, recommended the inclusion of a quantitative validation phase that includes one or more tests focused on content validity, criterion validity, reliability and construct validity. Additionally, in a scoping review of Øygarden et al[53] on measurement instruments for parental stress during the postpartum period, it was reported that none of the 15 instruments contained information on measurement error, responsiveness, and interpretability. Echevarría-Guanilo et al[54] argue that researchers should have a comprehensive knowledge of psychometric properties to tailor the research design to the most appropriate psychometric properties of the instrument of interest.

Regarding methodological approaches, Machado et al[43] identified the most widely used cross-cultural adaptation methods in nursing, and found studies where researchers added methodological approaches to the method they followed and studies where researchers did not comply with all the established methodological steps. Cruchinho et al[55] in a study that evaluated the methodological approaches used in the process of translation and cross-cultural adaptation of the Bedside Handover Attitudes and Behaviours (BHAB) questionnaire[56] reported the suplemental use of Dual-Focus to increase conceptual equivalence between the "source version" and the "target version". A methodological approach is defined as the way in which a phenomenon is studied systematically, shaped by the researchers' ontological and epistemological frameworks.[57] Applied to cross-cultural validation studies, it can be defined as a way of studying the equivalences intended to be achieved through the translation, adaptation, and validation of measurement instruments.

In the cross-cultural adaptation of instruments, different methodological approaches can be used for translation, such as: 1) one-away translation; 2) Dual-Panel approach, and 3) forward and back translation.[58] The one-away translation is the fastest and cheapest method, since it only includes bilingual individuals who translate the instrument into the "target language".[59] The forward and back translation is the most recommended method in translation guidelines.[21,60–62] It requires at least two independent translators: one translates the instrument into the "target language", and the other translates this version back into the "source language".[58] The Dual-Panel approach is a kind of Committee Approach involving a consensus translation by a panel of native bilinguals for the "target language", along with a member of the research team adapting the measurement instrument. This consensus version is then reviewed by a second panel of monolingual target population members.[63] It can also include a third panel to translate the translated version back into the "source language".[58] Lee et al[64] found that both the forward-backward and Dual-Panel methods enable the production of semantically equivalent translations and highlight that translation alone cannot eliminate cultural discrepancies.

Papadakis et al's[65] study comparing translations by translators with different characteristics, emphasized the importance of translators preferably being bicultural and having some content knowledge of the instruments, ideally selected from the target population. In-depth knowledge of everyday contexts (beliefs, values, habits, symbols, expressions) enables culture be reduced to a set of core variables for a given construct and facilitates cross-cultural research.[66] Members of the target population could be patients with literacy skills to enhance cross-cultural adaptation.[67] In addition, Papadakis et al[65] concluded that Principal Component Analysis of the measurement instruments is a methodology that can be used to compare translations carried out by translators with different profiles.

Methodological translation approaches can be symmetrical or asymmetrical. Symmetrical translations aim to make the instrument culturally relevant to the target population, while asymmetrical translations correspond to literal translations and maintain an one-to-one word correspondence.[58] In a study that found some confusion among translators about which approach to take when performing back translations, whether more asymmetrical and literal or more symmetrical and understandable in the "target culture", Bundgaard e Brøgger,[68] stated that guidelines provide specific instructions on the translation process and strategy. This was to ensure clarity of item meaning and minimize threats to construct validity. In order to facilitate the negotiations of committees of translators in relation to the nuances of items and consequently minimize threats to construct validity, other authors have suggested providing a description of intentions for each of the items.[69]

Cha et al[6] argues that the Committee Approach contributes to acceptable internal consistency coefficients. Concurrently, Epstein et al[70] found that carrying out a multidisciplinary expert committee contributes to obtaining rigorous items in the adaptation of a multidimensional instrument.[70] Other authors have reinforced the relevance of different types of Committee Approach. For instance, Teig et al[71] reported that using the Delphi method in an Expert Committee with the criteria of anonymity, controlled feedback and statistical responses, provides a more accurate measure of the degree of consensus of all the elements than if a meeting had been held without any formal voting system.[71] Tsai-T-I[72] described a process of cross-cultural adaptation that involved a panel of experts to determine the content validity of the original instrument before translating it into the target language. Also, Jayawickreme et al[73] stated the importance of using a Focus Group series to promote the evaluation of translated items by a panel of experts.[73]

Montenegro et al[74] highlighted the importance of using Dual-Focus as a decentering strategy in the context of the Committee Approach. After forward-backward translation, items or parts of items that are not appropriate for the "target culture" may be identified. In these situations, decentering and Dual Focus can be used.[75] Decentering is a translation procedure that does not require a literal translation, which is used to achieve idiomatic, grammatical-syntactical, experiential and conceptual equivalence between the two cultures.[6] Dual-Focus involves replacing items or parts of items with more appropriate ones in the "target language" in order to mitigate the difficulty of adapting certain content from the "source culture".[22] It allows us to scrutinize what each of the items in the "original version" of the instrument seeks to assess in the light of the operational definition of the construct we want to measure, and thus, ensure that we are concerned with content validity.[76] Several studies have reported the substitution of words and items as a result of using Dual-Focus.[77–83]

The specific relevance of different methodological approaches has been justified in scientific literature. Toma et al[84] highlighted the effect of combining the back translations with Cognitive Testing (also called Cognitive Interviewing and Cognitive Debriefing) in modifying five items of an instrument with each of these approaches.[84] Comparing the results of the Cognitive Debriefing with the original instrument is essential to ensure cultural relevance, since it can reveal problems with wording, phrasing and resonance with individual's world views.[85] Hasani et al[86] recommended the inclusion of Cognitive Debriefing in the research design together with the Expert Committee approach to ensure the validity and reliability of the measurement model.[86] However, in an integrative literature review which analyzed how back translators were described in 105 empirical studies, Bundgaard e Brøgger[87] found limited information on translators' qualifications in empirical studies.

Back translation is the methodological approach whose importance has been justified in various ways. It was first advocated to limit the substitution of item content for cultural reasons.[88] Subsequently, other researchers defended its use not as a method of equivalence, but rather as a way of checking the content of the items and the purpose of the instrument.[89] More recently, the use of back translation has been argued as a documentation tool to show "what the translation says" and thus support researchers' decision-making when adapting the instrument.[90] Epstein et al[70] concluded that back translation has little effect on the content and psychometric properties of a multidimensional instrument. Despite this, the same authors warned that back translation is an essential methodological approach for the authors of measurement instruments when they are not proficient in the "target language".[70,87]

Previously published guidelines and recommendations present a set of methodological approaches in a prescriptive way,[21,60,62,91,92] that does not promote the researcher's decision-making on the best options for the characteristics of their validation studies. To facilitate the decision-making process several authors have been proposed glossaries,[93] decision trees,[94] and checklists,[95–98] which enable researchers to avoid gaps in the process that affect the quality of the final instrument,[52,99] and promote the active role of researchers in conducting the processes. The scarcity studies comparing methodological approaches prevents the recommendation of a specific method,[25,44] meaning the processes of translation, adaptation and cross-cultural validation dependent on skills, knowledge and time,[100] something that young researchers may not always have. Peña[27] and Arafat[101] recommended developing guidelines to support researchers' decisions throughout the process. Similary, Cruchinho et al[55] called for comprehensive guidelines on methodological approaches for novice researchers decision-making. Comprehensive knowledge of methodological approaches is a prerequisite for cross-cultural validation studies of measurement instruments.

Research Rationale and Aims

The first guidelines produced for the healthcare field emerged from extensive literature reviews, including literature from the health, psychology, and sociology.[21,42,102] In Brazil, an integrative review of nursing literature revealed an over-emphasis on evaluating psychometric properties at the expense of exploring methodological approaches for translation and cross-cultural adaptation.[41] To date, no study has identified the differences and similarities between existing guidelines in healthcare to support young researchers in the development of validation studies. Based on this, we formulated the following review question: - What similarities and differences exist in the methodological approaches recommended by existing guidelines on the process of translation, adaptation, and cross-cultural validation of measurement instruments in healthcare sciences? Therefore, this study aims to: 1) identify all guidelines and recommendations for translation, cross-cultural adaptation, and validation within the healthcare sciences; 2) describe the methodological approaches established in these guidelines, and 3) provide a practical guideline featuring various methodological options for novice researchers involved in translating, adapting, and validating measurement instruments. If you are planning to translate or adapt an measurement instrument, this article will assist you in critically choosing a methodological approach to obtain a valid, reliable, and unbiased instrument.

Materials and Methods

Identification of Existing Guidelines

A methodological review was undertaken for this study. Methodological review is a type of literature review focused on summarizing the state-of-the-art in methodological practices within a particular domain.[103] In this methodological review, the focus was on the methodological approaches used for the translation, adaptation, and cross-cultural validation of measurement instruments recommended by guidelines in the field of healthcare sciences. For this methodological review, we used a three-stage search strategy.[104] The initial search was limited to the "CINAHL with Full Text" (via EBSCO) and "MEDLINE with Full Text" databases and included an analysis of the text words in the titles, abstracts and indexed terms used to describe the manuscripts in each of these databases. The second search involved the Boolean expression (((MM "Instrument Adaptation") OR "cross-cultural translation" OR "cross-cultural validation") AND ("recommendations" OR "best practice")) in the "CINAHL with Full Text" (via EBSCO), and "cross-cultural adaptation"[Title] OR "cross-cultural translation"[Title] OR "cross-cultural validation"[Title]) AND ("recommendation*"[Title/abstract] OR "best practice*"[Title/abstract] OR "methodological approach*"[Title/abstract])) in the "MEDLINE with Full Text". Finally, we reviewed the reference lists of the manuscripts obtained to identify any guideline(s) that were not retrieved in the initial literature search in the databases. We used this search strategy because it allowed us to identify additional guidelines. The database search was carried out between September and October 2023 and it was repeated in February 2024 to capture any guidelines that had been subsequently published. To select the articles, our criteria were based on the concept that a guideline summarizes evidence and expert opinions, considering existing resources and the feasibility of procedures.[105] The inclusion criteria for selecting the manuscripts were: 1) a scientific article focused on the process of translation, adaptation, or cross-cultural validation of measurement instruments; 2) describing a guideline to be followed in one of these processes; 3) written in English, Spanish, or Portuguese; and 4) published in a scientific journal in the field of healthcare sciences. To define the areas of healthcare, we used the Classification of Health Care Providers (ICHA-HP) framework,[106] which describes the actors who provide health care (eg general and specialized physicians, nurses and midwives, physiotherapists and physical therapists, occupational and speech therapists, audiologists, dental hygienists, mental health specialists, etc.). The exclusion criteria for manuscripts were: 1) editorial articles, literature reviews, thesis, dissertations, or book chapters, or scientific articles not focused on the process of translation, adaptation or cross-cultural validation of measurement instruments; 2) articles that do not describe any guidelines to be followed in one of these processes; 3) guidelines that have already been included for eligibility; 4) articles written in a language other than those specified; and 5) articles published in a field other than healthcare sciences (eg economics and management, education science and sociology). Figure 1 shows the results of the search and the selection of studies.

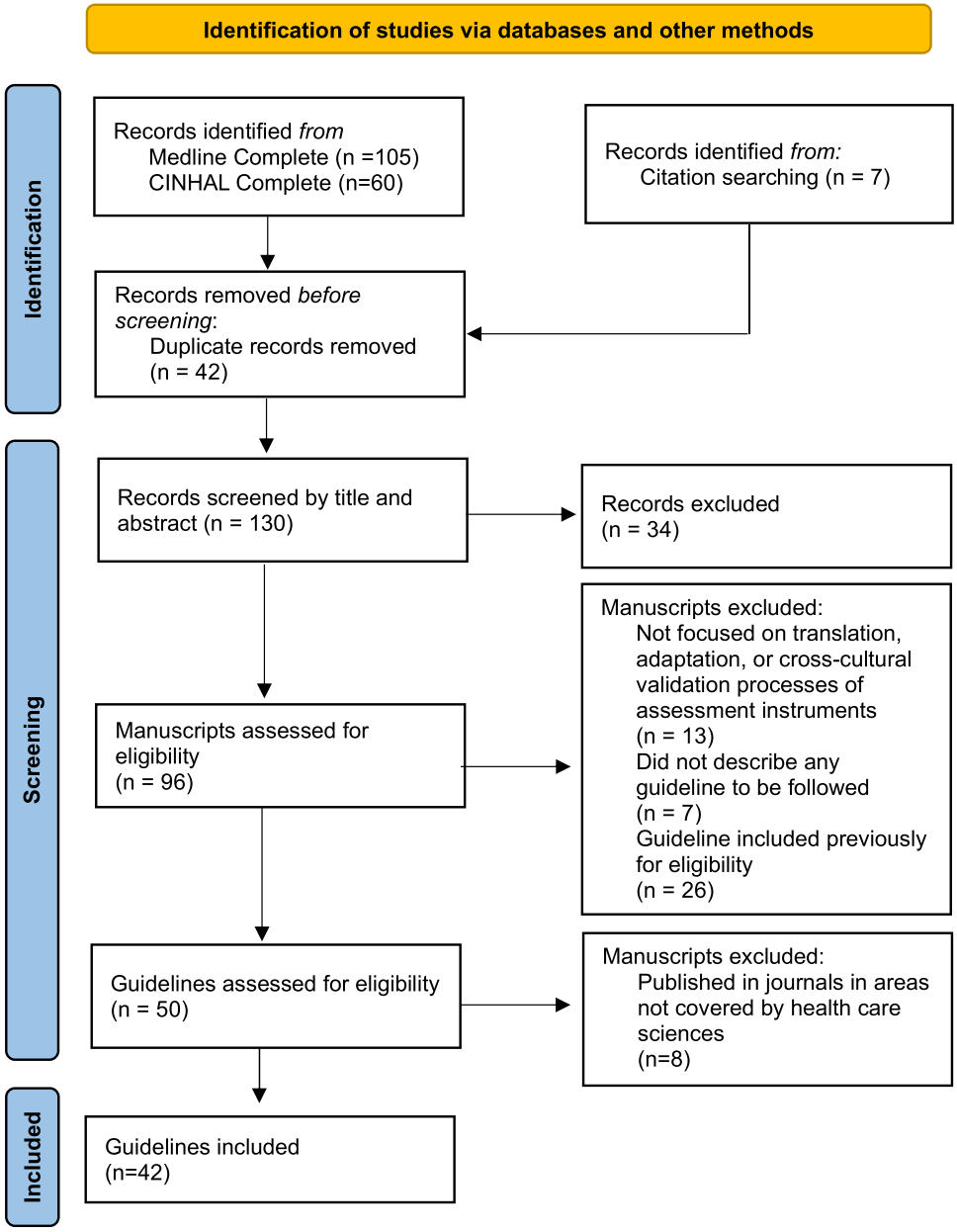

Figure 1 PRISMA Flow Chart of literature review. Adapted from Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021:n71. Creative Commons.[107]

Content Analysis of Existing Guidelines

A content analysis was conducted,[108] which included a total of 42 guidelines retrieved from the literature search. Based on this, we established two objectives: 1) to provide an overview of the range of methodological approaches included in the guidelines and 2) to identify the similarities and differences that exist in terms of the methodological approaches recommended. In a first step, all relevant excerpts from the guidelines focusing on methodological approaches were paraphrased, summarized, and structured. Based on these excerpts, paraphrases were formed, and categories were inductively generated. Subsequently, the categories generated were reviewed and grouped by similarities and differences into broader thematic categories. Finally, the paraphrases and categories derived from the guidelines were described narratively.

Proposal for a Practical Guideline

The development of guidelines is a multidisciplinary process that should include all relevant areas of expertise and perspectives.[109] Based on the synthesis of the methodological approaches from existing guidelines, we have drawn up

a practical guideline based from an universalist perspective,[26] enriched by contributions from experts in the fields of nursing management, statistics, and linguistics with experience in the translation and cross-cultural adaptation of health measurement instruments and in the supervision of novice researchers. Our recommendations are grounded in the common elements identified among the guidelines retrieved from the methodological review, and are supplemented by our own professional expertise.

Results

We reviewed 42 guidelines on the processes of translation, adaptation, and cross-cultural validation of measurement instruments. The guidelines included were published between 1993–2021, and most were published during the first two decades of the 21st century (Figure 2). The main countries of publication were UK (7), Netherlands (7), USA (4), Canada (4), Spain (4), and Brazil (2) (Figure 3). The findings will be presented using the four thematic categories developed because of the analysis: general information, cross cultural translation, cross-cultural adaptation, and cross-cultural validation. Here, in keeping with the aims of this methodological review, we provide an overview of the similarities and differences of methodological approaches recommended by the guidelines about each thematic category.

General Information

Some guidelines recommend a preliminary stage before the translation of the instrument called preparation.[60,92,98] This stage includes obtaining permission to use the instrument,[60,92,96] without clarifying whether this permission is given by the authors of the instrument or by the publisher who holds the copyright to the articles. Other authors suggest that permission should be requested from the instrument's publisher,[98] the affiliation institution,[98] or the authors.[60,92,98,110] In addition, other authors indicate that permission should be obtained from the owner of the instrument's intellectual property rights.[91,99,111]

The initial phase also involves deciding which instrument to adapt cross-culturally. Some guidelines recommend that this decision should be based on checking that no version of the instrument exists for the target population,[98,112] understanding their context,[112] its purpose,[91,98,112] features,[112] the dimensions of the construct,[98] the conceptual equivalence of the construct for the target population,[91,98] its suitability for the intended clinical context,[91,112,113] adequacy of psychometric properties,[91,98,112] the existence of other cross-culturally adapted versions,[114] and feasibility.[98] Regarding feasibility, some authors specify factors such as completion time, cost and duration of the instrument, and the type and ease of administration.[115] Other guidelines recommend only identifying the evidence on the quality of the selected instrument.[92,116]

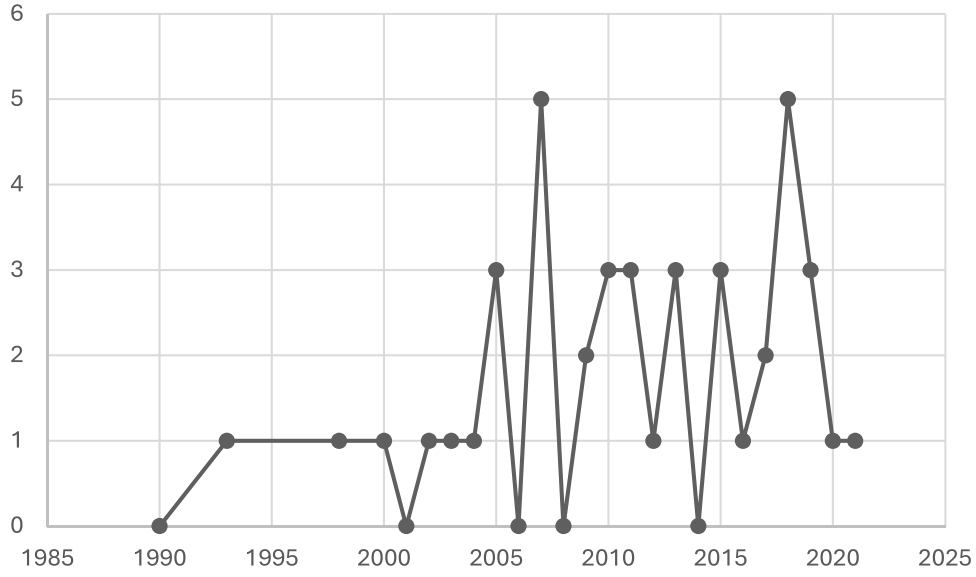

Figure 2 Distribution of guidelines by years.

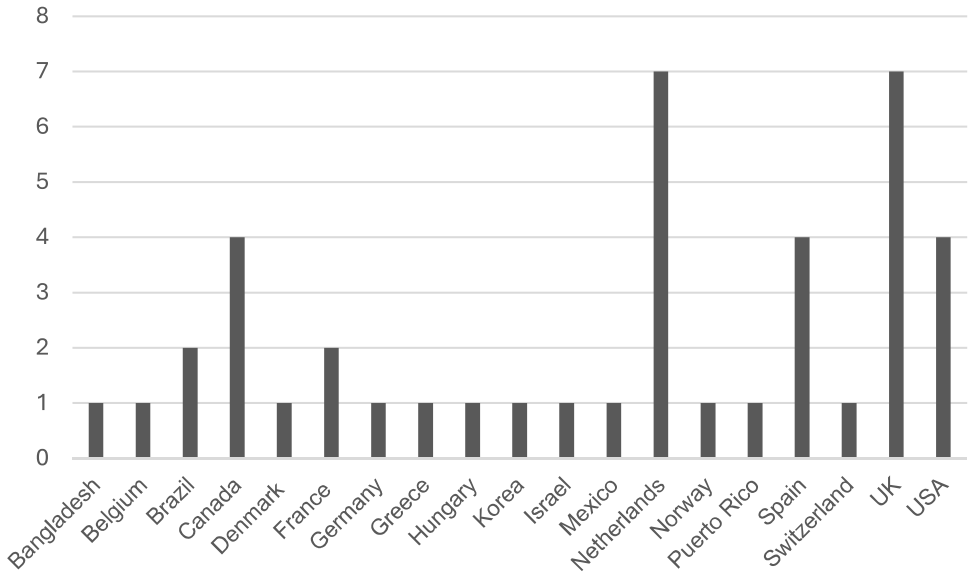

Figure 3 Distribution of guidelines by countries.

To support the decision on which instrument to adapt, some authors recommend studying the relevance of the construct in the target population,[99] as well as its conceptual framework,[116] or meaning,[96,116] which can be carried out through a literature review,[26,61,117,118] open interview or Focus Group,[26,61] or observation of members of the target population.[9,26] Other authors emphasize the importance of researchers identifying early on the cultural and linguistic differences between the "target culture" and the "source culture".[92,96,111] To facilitate the translation and adaptation process, some authors propose developing a definition of the instrument's constructs.[98,112,116] Some guidelines recommend providing translators with information about the instrument's construct,[26] that can help translators resolve cultural and linguistic differences between the "source culture" and the "target culture",[119] eg scientific articles.[91]

Also included in the preparation phase is the design of a protocol for the process of cross-cultural adaptation of the selected instrument.[91,99] Some authors recommend researchers decide which method to use:1) the same language adaptation approach for instruments adapted in another country or population with the same language; 2) the universal approach for translations intended for multiple locations simultaneously; or 3) the country-specific approach for different translation versions developed for each subpopulation.[120] The use of combined methodological approaches and procedures that maximize conceptual equivalence between the translated version and the "original version" is proposed by some authors,[111] which may vary according to the particular characteristics of the studies and the resources available to the researchers.[121] This includes setting up a multiprofessional team comprising translators and experts in the field of the instrument's construct.[96]

Few guidelines specify the leadership role of researchers within the team of translators and experts, for example in reviewing decisions to reconcile translations and in producing a more literal or more conceptual translation.[60,98] Some special roles can be assumed by researchers, such as qualified moderators of Expert Committees,[116] translation coordinators,[122] or reviewers[6,98,112,120,123] of proofreading after forward translation,[6,98,112,120] back translation,[92,120,122,123] pre-tests,[60] and the final version of the adapted instrument.[98] Some authors also advocate including members of the target population in the team when reviewing the cultural differences of the assessment instrument,[124] and its developers,[26] in the translation and cross-cultural adaptation process,[26,98] and the study of the instrument's measurement properties.[116]

Many guidelines recommend documenting the translation process, cross-cultural adaptation, and validation in a report that describes all methodological approaches and procedures used, and their results, problems identified, proposed modifications,[21,60,92,93,96,98,99,112,116,121–123,125–127] along with the names, roles and background of all those involved,[125] the testing process, and the statistical analysis.[128] Some authors recommend creating a template for continuous recording of the process.[98,112] The information in this template can be used to prove the equivalence between the adapted and original versions

of the instrument and as supplementary material in the publication of a scientific article reporting on the process of cross-cultural adaptation of the measurement instrument.[121] McKenna,[114] argues that the overall process should be reported.

Cross-Cultural Translation

Forward translation and subsequent back translation is recommended in some guidelines.[6,9,21,27,42,60–62,91–93,98,110,116–120,122–127,129,130] Some authors propose forward translation without back translation,[99,114,118,128,131] or suggest back translation as an option.[112] The majority, also propose that forward translation and back translation be carried out independently.[9,42,60–62,80,91–93,98,99,110,112,116–119,122,123,125–127,131] Despite this, some authors propose the use of collaborative approaches to translation, such as the Committee or Focus Group Approach,[128] the Dual-Panel,[114] the use of a Bilingual Committee,[118] translation with teams of two translators,[80,129] or two to four translators.[91] Others authors recommend using the one way or expert´s translation by a committee when human or financial resources do not allow the back translation to be planned.[128]

Regarding the number of translators, most guidelines suggest using at least two translators to translate the instrument and the same number for the back translation,[9,21,42,60–62,92,93,98,110,117,122] or two translators for each of these approaches.[9,91,123,126] Others propose at least two translators for forward translation and at least one for back translation,[60,92,98,129,130] and others, at least one different translator for forward translation and back translation.[27,118,128] Other authors recommend three translators for forward translation and one for back translation.[6] Some of the guidelines that do not recommend back translation recommend using one translator,[118] and two translators for the forward translation.[112,131] Other authors recommend multiple translations without specifying a minimum number.[99,116]

Some guidelines recommend that researchers give translators instructions on which translation approach to follow (whether more literal or more cultural),[119,125,131] explaining the concepts of the measurement instrument and how to use its definitions in each of the items,[60,98] which may involve the prior supply of materials.[91,98] Others only recommend providing information about the purpose of the instrument, the target population and the aim of the translation if the study involves professional translators.[92] Most authors recommend using native translators,[21,26,60,61,80,91,92,110–112,116,118,119,122,123,126,127] who are bilingual in both the source and target languages.[6,21,26,62,91–93,98,110,111,117,123,125,128] Other authors suggest that the translation process should include at least one different professional translator in both forward and back translation approaches,[92,98] others establish only the inclusion of professional translators in both approaches,[96,127] and others only for back translation.[93]

In relation to the translators involved in forward translation, various characteristics are described, for example: 1) familiarity with the construct of the instrument;[6,9,62,92,99,111,112,116,123,125] 2) be a health professional familiar with the terminology used in the measurement instrument,[91,112,119] or with experience in the clinical condition of interest;[98] 3) have translation experience;[93,112] 4) have previous experience of PROMs,[60] and 5) be a representative member of the target population.[98,128] Some authors require having a translator familiar with the instrument's construct and an unfamiliar translator,[21,62,110,116,123,128] while others require two translators familiar with the instrument and the context.[6,9,42,91,92,125,126] It is, also, recommended that translators taking part in back translation: 1) not have access to the "original instrument",[21,62,92,98,110,123–125,127] 2) are both naive about the construct to be measured,[21,42,92,98,116,119] or one of them familiar with the construct area of the instrument and the other familiar with the linguistic and cultural nuances of the "source language".[62]

Most guidelines establish a synthesis or reconciliation stage after forward translation to identify and resolve discrepancies in the translation.[21,60,62,93,98,110,119,123,131] To do this, some authors recommended using a third translator to reconcile the two versions of the forward translation,[61,62,98,117,131] or one reviewer,[60,98] or two.[6,126] To obtain more accurate translations, researchers, also can provide materials to reconciliation translators.[119] Some guidelines stipulate that the reviewer should reconcile the two translations into a single version together with the translators involved,[91] including elements of the target population.[61] To discuss and reach consensus on the differences found between forward translation versions, can be used the Committee Approach,[6,130] Focus Group,[61] or Delphi Panel.[92] If a consensus cannot be reached on some of the discrepancies, one of the developers of the "original instrument" may be involved,[91] or the use of an independent translator, who decides on the translation alone with input from the developers of the measurement instrument or another forward translator.[98,122] Others recommend using the Delphi Panel to evaluate and resolve discrepancies between

the constructs with elements not involved in the previous translations but with extensive knowledge of the constructs to be measured.[92] As an alternative to using a collaborative approach, some authors propose having the back translated version reviewed by an independent translator who decides on semantic equivalence by comparison with the "original version" of the measurement instrument.[6,123,127] Other authors suggest that the forward and back translation versions be synthesized by the same independent translator.[117]

With regard to deciding on the differences found, some authors argue that the meaning of an original term can be modified during the translation process if only part of the meaning is present in the "target culture" or if the term in the "target culture" expands the meaning of the term in the "source culture".[125] Some authors establish criteria based on source and comprehensibility, cultural appropriateness, grammar and terminology to support the reconciliation process decision.[122]

Cross-Cultural Adaptation

To adapt instruments cross-culturally, several guidelines include the use of collaborative approaches in the form of Expert Committees.[9,21,42,62,92,93,96,98,99,116,117,123] For some authors,[117] the committee of experts aim to ensure only semantic equivalence, while for others additionally aim to ensure idiomatic, experiential equivalence,[21] and conceptual equivalence.[21,62] Authors with a narrower purpose call the committee the Bilingual Committee,[118] and the authors with a broader purpose call the procedure carried out by the committee of experts the Harmonization procedure,[60,92] and the committee, the Review Committee,[9,42] or Multiprofessional Committee.[21,62,93,98,117,132] The latter includes translators, linguists, methodologists and psychometricians.[21,93,132] It may also include a monolingual element.[62,123] Some authors also include members of the target population,[60,62] who are preferably independent of the project team,[98] who can be health professionals.[93] Other authors promote holding a meeting with experts from the target population after the Expert Committee.[61,117] The developer of the measument instrument may also participate if he or she is proficient in the "target language" or can be contacted to clarify any issues.[62,98,117] In the denomination of the committee of experts, some authors have adopted the term "professionals" instead of "experts" because they consider that, in relation to PROMs, these individuals are the main experts.[133]

Contact with the developers of the "original instruments" is advocated by some authors,[21,92,98] especially when researchers want to eliminate items before psychometric analysis,[132] or when omissions are identified in the measurement instrument.[132] Regarding the adaptation of items, some authors advocate that items can be adapted to maintain their meaning when: 1) literal translation into the "target language" is not possible due to the lack of words or 2) the items in the "source language" include idiomatic expressions.[121] The agreement of the instrument's developers is required whenever parts of items need to be replaced,[91] or of complete items.[121] Some authors recommend avoiding the inclusion of new items and the elimination of parts of items or complete items prior to psychometric validation.[121,132] Others consider the possibility of eliminating items, as long as they are items of low cultural relevance.[123] The Decentering or Dual-Focus techniques are recommended for adapting items.[27]

Most guidelines include pre-tests to cross-culturally adapt measurement instruments.[6,9,21,26,27,42,60–62,92,93,96,98,99,110–112,114,116–119,122–124,126–130,133] For some authors, the pre-test involves conducting cognitive debriefings with members of the target population,[60,116,127] using Focus Group[98,118] or a Delphi Panel,[26,133] to evaluate the interpretation of items in the harmonized version and to identify wording that may be unclear. Cognitive debriefing interviews consist of: 1) asking participants to answer the questionnaire;[128] 2) paraphrase the participants' understanding, item by item, in order to identify the items that may have translation problems;[127,128] 3) asking if they would write the items differently, how they selected their answers, if they identified any words they did not understand and if they considered any expressions unacceptable or offensive,[119,127] and 4) asking about relevant topics that could be included in the questionnaire.[128] There is no consensus on the number of participants in the group debriefing, which can involve between three and 10 participants,[127] five to eight elements,[60] at least eight,[98,123] at least 10 elements,[110] or with 10 to 15 participants from the target population.[122,126] Respondents must be representative of the target population (eg in terms of gender, age, education and diagnosis).[60] Other authors recommend the possibility of this debriefing being video or tape recorded.[123] Several guidelines specify a cognitive debriefing with individual face-to-face interviews carried out with members of the target population,[61,98,114,119,122,124,126,128] and preferably recorded.[93] Also, there is no consensus on individual debriefings either, which can include five to eight respondents,[60,92] at least seven patients or seven health professionals representing the target population,[133] a sample of 30 elements,[129] 30 to 40 participants,[21,61,117,123] and 25 to 75 respondents.[112] Some guidelines recommend collecting sociodemographic information

from pre-test participants,[119] the recording and transcription of cognitive interviews,[116,133] and others the coding of the interviews by two independent researchers.[133]

The qualities of the measurement instruments to be tested recommended by the authors are diverse. A clarity pre-test is included in several guidelines,[9,62,96,99,110,119,130,133] as well as a pre-test of comprehensiveness,[9,26,61,80,96,110,117,119,130,133] of acceptability,[9,61,117,124] the relevance of the items,[26,62,99,110,116,124,130] and the emotional impact of items.[61,117] Some authors recommend assessing comprehensiveness and relevance separately, followed by a cognitive interview to assess clarity.[133] Other authors suggest evaluating the coherence and comprehensiveness of the items, the operational aspects, along with relevance and clarity.[110] Others propose evaluating clarity, comprehensiveness and acceptability.[112] In relation to the PROMs, it is proposed to carry out pretests with patients to assess the relevance, the comprehensiveness, and the comprehensibility and with health professionals to evaluate only the relevance, and the comprehensiveness.[116,133] Some authors include in the pre-test of comprehensiveness, evaluating the time needed to complete the instrument,[99] by participants or researchers.[130] Before administering the pretests, a critical review of the adapted instrument by members of the target population may also be included.[112]

In the pre-tests evaluating clarity, relevance, comprehensiveness, coherence and operational aspects, it is suggested using a visual analog scale or a Likert-type scale to assess the content validity of the adapted instrument.[110] To assess this type of validity, it is recommended to calculate the Content Validity Index (CV-I) of the measurement instrument and the CV-I of each item,[62,110,123,124] and the Kappa Coefficient Agreement.[62] Items with unacceptable values are reviewed and reevaluated.[62] Some authors state that in conjunction with the assessment of content validity, also may be carried out a statistical analysis (eg, Rasch item analysis and Cronbach's α).[124] Keeping the adapted instrument with the same format of items and response options as the "original instrument" is recommended by some authors.[96] Other authors recommend that the research team discuss the format, instructions, mode of administration and measurement methods used by the "original instrument" with the members of the target population.[61,117] After the pre-test, some authors specify the need to decide on the form of dissemination of the instrument (whether through a paper questionnaire or an electronic questionnaire).[61]

Cross-Cultural Validation

Many guidelines do not include psychometric validation as a step in the cross-cultural adaptation process.[9,27,60,92,98,119,122,126,129] Others only include psychometric validation without covering the translation and cross-cultural adaptation stages.[102,115,134–136] Others provide information on the psychometric properties to be evaluated,[6,9,21,61,91,99,110–112,114,118,120,121,123–125,128,130,133,137] and others detail information on statistical procedures and analysis methods for evaluating certain properties.[26,62,93,102,116,134,135,138]

With regard to statistical procedures and methods, some authors include information on sample requirements for psychometric validation, namely that the number of participants should be taken into account on the basis of the number of missing values,[116] the power of statistical testing,[124] or that saturation is more important than sample size.[133] Others specifically propose getting > 100 participants as a very good criterion for assessing internal consistency, measuring error and reliability, testing hypotheses for construct validity and comparing subgroups.[116] Other authors propose a sample size of 100 and 200 respondents,[112] and another at least 200 participants.[111] Others indicate a ratio of 10 participants for each item in the instrument.[62] Convenience sampling is recommended for sample selection,[124] with characteristics relevant to the intended use of the instrument.[96,111]

Regarding psychometric properties, we found a wide range of information. Some authors propose evaluating the reliability,[6,9,21,61,99,112,116,124,136] while others specify the evaluation of the Cronbach's α,[26,93,121,127,128,130,138] the Interclass Correlation Coefficient (ICC),[26,93,110,121,130,138] or K-index,[93,110,130,138] as an indicator of test-retest reliability. To assess this property, it is recommended to apply the same adapted instrument to the same respondents at seven and 14 days.[110] To assess internal consistency, several authors recommend using the Exploratory Factor Analysis (EFA) followed by Confirmatory Factor Analysis (CFA),[61,62,121,132] or Exploratory Structural Equation Modeling as an alternative to confirmatory factor analysis.[121,132] Others also include item-total correlation, inter-item correlation and Differential Item Functioning (DIF).[110,128,139] The purpose of the DIF assessment is to compare the level of an item between two different groups of different levels using the same instrument or to identify items that may cause measurement bias.[128] To assess cross-cultural validity,[116,137] EFA followed by CFA is also recommended.[134] Regarding the validity of the instrument, some

authors refer generically to its evaluation.[9,112,130] Others specify the measurement of content validity,[21,50,93,125,134,136,138] and construct validity,[21,26,93,99,116,118,121,124,125,128,134,136,138] particularly by analyzing the factor structure of the instrument (dimensionality).[6,26,61,62,136] Criterion validity is another recommended property,[6,116,118,134,136,138] in particular discriminant validity,[26,62,110,121,123,134] predictive validity,[61,62,93,110,121] convergent validity,[26,62,110,121,123,125,134] and concurrent validity.[61,62,93,110,121,134] Some authors state that two or more instruments can be validated concurrently, especially when there are known relationships between their constructs.[132] Finally, other authors propose the evaluation of measurement error,[62,110,116,136,138] responsiveness,[21,26,110,115,116,124,134,138] floor and ceiling effects,[133] and hypothesis testing.[136]

Based on the psychometric properties, some authors suggest evaluating the quality of the evidence for each measurement property, namely content validity (with evidence that the instrument's items are relevant, clear and understandable in relation to the construct of interest and the population being studied), structural validity (with evidence of the factor analysis or Item Response Theory (IRT)/Rasch analysis), internal consistency (Cronbach's α), cross-cultural/measurement invariance with evidence from DIF or Multigroup Confirmatory Factor Analysis, and the remaining measurement properties (reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness).[115] Based on this evidence, some authors propose that researchers assess the degree of item and semantic equivalence, operational equivalence, functional equivalence and conceptual equivalence and measurement equivalence.[135] For example, if factor analysis reveals structural differences between cultures, it is advisable to assume that the instrument is not equivalent between cultures. In such cases, it is recommended to resort to qualitative research methods to understand the reasons behind this lack of equivalence.[135] In addition, some authors recommend comparing the psychometric properties obtained with those reported by the authors of the "original instrument",[21] and reviewing the adequacy of the psychometric properties with the team,[124] as well as with experts in the field of the instrument's construct and members of the target population.[6,91]

Discussion

The use of a rigorous methodological approach helps minimizing the occurrence of biases during the process of translation, adaptation, and cross-cultural validation of measurement instruments. Although there is currently a wealth of guidelines in the literature, researchers often focus solely on the translation aspects and do not use them as a methodological guide in their studies.[62] Some of this information has been disseminated as standards by some scientific organizations. For example, the International Test Commission has disseminated, in several languages, standards related to the decision of adapting a measurement instrument, the translation and adaptation process, the empirical validation process, the scoring and interpretation and the documentation of the procedures used.[111] Additionally, in USA, the American Educational Research Association established a set of standards for assessing validity and reliability that includes characteristics of the design and development of measurement instruments.[140] Several researchers across different fields have published models, guidelines, and detailed recommendations for the processes of translation, adaptation and cross-cultural validation.[21,26,60,62,91,141,142] These guidelines vary in terms of prerequisites, the number of stages, the number and profiles of the translators involved, the configuration of the experts panel and their profiles, the inclusion of reviewers of the translations, methods of identifying bias, and the number and method of conducting pre-tests. In some scientific areas, there might even be a preference for using a single methodological approach. However, it is not mandatory to follow all the guidelines steps and procedures in the instrument validation once the guidelines may not be applicable to all studies' characteristics.[143] Even when researchers choose to follow a specific guideline to develop the process, this does not preclude the possibility of customizing the methodological approaches prescribed at each stage. Such customization can be justified on the types of equivalence to be achieved or reinforced and on the biases to be avoided.

Process Documentation

The identification of reasons for the different functioning of an adapted or validated version of a given measurement instrument can lead researchers to consider the possibility of bias during the process of translation, adaptation and cross-cultural validation.[22] To check this possibility, researchers need to ensure that the process is traceable with records in each stage.[21] Those records allow, for example, that researchers verify that the low relevance of a particular item is not associated with any type of bias and may, as a result, be modified or excluded before the psychometric validation of the measurement instrument. At each stage, researchers involved in the process of adapting measurement instruments need to provide formal

written evidence of the probable relevance of the instruments to participants from the "target culture", as well as of the operational equivalence of the instrument. The documentation of decisions may be empirical but most will be theoretical in nature.[144] That documentation includes, by instance, information about the different stages of the process and the activities carried out at each one. It should also include information about the decisions, rationale and reasoning behind those decisions, as well as the professionals who participated in the different activities.[145] Documentation of partial modification or removal of items is one of the main problems in the cultural adaptation of measurement instruments and requires adequate justification.[146] These information must enable other researchers to understand and evaluate the work carried out,[147] and to replicate the used procedures both in the same population and in others.[111] Transferability is a key concept when a specific measurement instrument is used in different cultures and contexts. Considerations about the transferability of a measurement instrument are supported by documentation on the relevance of the construct, the measurement method used, the translation strategies adopted, and the cultural practices that may influence the results.[20] Cruchinho et al[55] made available the different versions of the cross-cultural adaptation process in supplementary material with the publication of the translation and cross-cultural adaptation process of the BHAB questionnaire.[56] After writing your study protocol, be sure to create a documentation model in which you will record all the steps and all the methodological approaches to be used and their results.

Stages of the Process

Performing the field testing of the measurement instrument separates the translation, adaptation and cross-cultural validation procedures from the psychometric evaluation procedures which determine the global analysis of its properties. On this basis, we organized the overall process into eight sequential steps with distinct purposes, which are: 1) forward translation; 2) forward translation synthesis; 3) back translation; 4) harmonization; 5) pre-testing; 6) field testing; 7) psychometric validation and 8) analysis of the psychometric properties (Figure 4). Following, we will describe each one of these steps.

Forward Translation

Translation from one language to another is not always straightforward due to the variety of possibilities for translating a word or expression, and because there may not be an equivalent word in the "target language" to represent a particular term.[91] Consequently, different translators may choose different translations options for the same instrument that do not coincide. To achieve a clear translation, Brislin, Lonner and Thorndike,[148] produced a set of 12 guidelines. These recommendations establish the need for the items to: 1) include short, clear and simplified sentences; 2) use the active voice rather than the passive voice; 3) use nouns repeatedly rather than pronouns; 4) avoid the use of metaphors, regional phrases, idioms or colloquialisms; 5) not use the conjunctive mode of verb tenses; 6) use additional phrases to ensure understanding of item content; 7) not use items that include adverbs and prepositions; 8) avoid item content that includes possessive forms of words; 9) be specific; 10) not use vague descriptors; 11) familiarize the translator with item content, and 12) avoid more than one verb for item content that suggests different actions. The concern with the clarity of the items makes it possible to guarantee conceptual equivalence between the target and source versions. For this procedure, most methodological approaches propose a minimum of two different translators.[21,62,91,149] Regardless of the number of translators, it is crucial that the translations are performed independently, that translators do not discuss the translation before completing it, and that their work is not affected by the knowledge that the translation will be subject to back translation. Ozolins,[150] argues that if forward translators are aware that a document will be back translated, sometimes they can opt for more literal terms rather than choosing terms that are culturally appropriate in the "target language."

Using professional translators working in the language pair in question allows comparisons of the versions of the measurement instrument in both languages, facilitating the identification of ambiguities and discrepancies, even when there is semantic equivalence. One of the main requirements is the familiarity with both cultures to be able to recognize situations and items for which a literal translation may be inadequate. Forward translators should also make suggestions for items in the culture in which the instrument is to be adapted, even if the item is left with a different meaning from the original.[22] For example, translators can suggest replacing the terms "nursing assistants" and "advanced practice nurses", respectively, by "auxiliary staff", and "specialist nurses", in the countries where these categories do not exist. In addition to being bicultural, some methodological approaches recommend only using professional translators.[60] Others define the inclusion of translators who are native speakers of the "target language",[111] and have a specific profile, for example, being knowledgeable about the

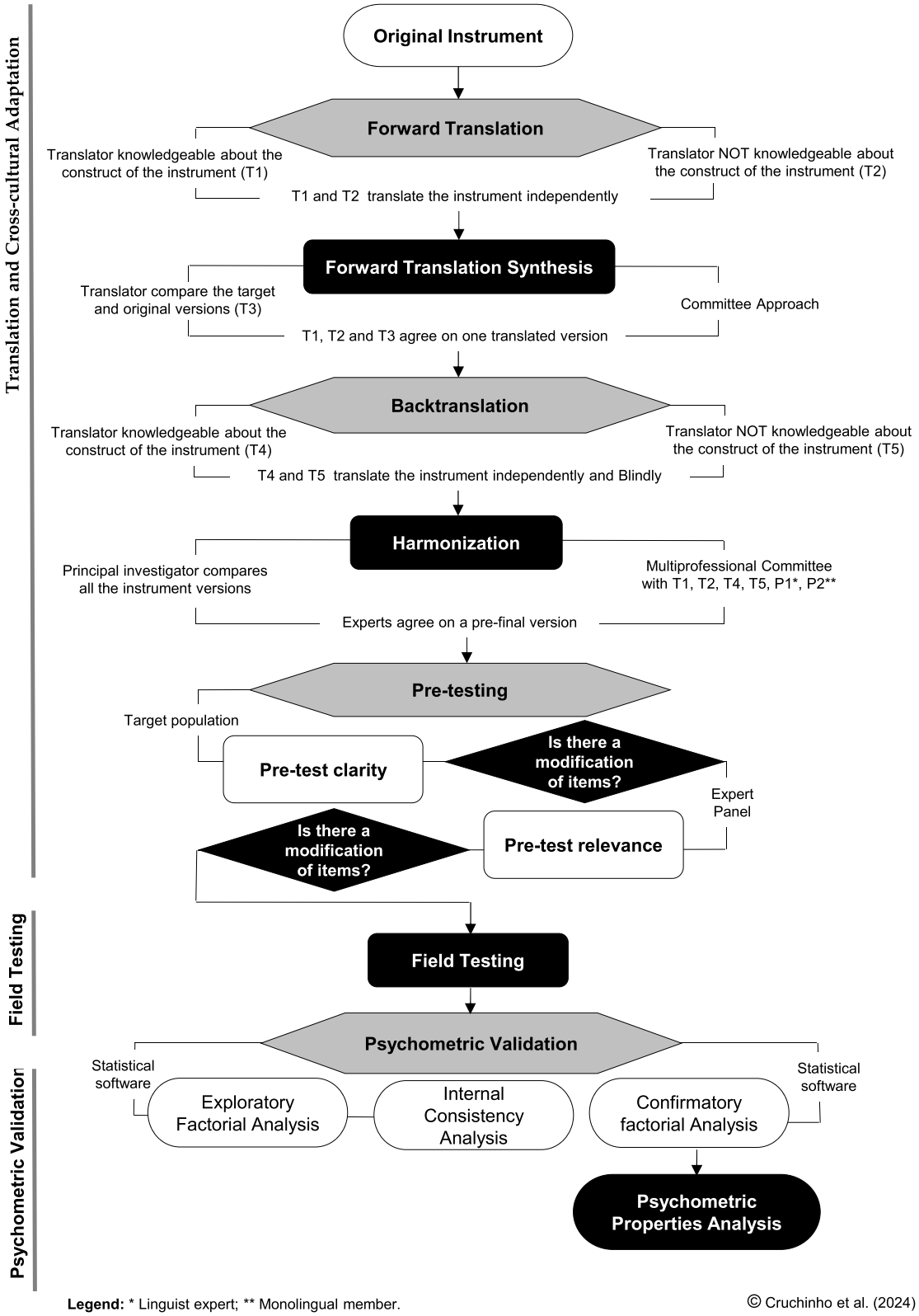

Figure 4 Flowchart of the translation, adaptation and cross-cultural validation process.

construct of the instrument.[21,62,111,149] To minimize the risk of content bias, members of the target population in which the measurement instrument is intended to be applied, may be involved in the translation.[151] If members of the target population cannot be involved, members acquainted with the terminology used by the measurement instrument can be chosen.[21,62] According to Hedrik,[22] the translation of measurement instruments does not aim to obtain a precise translation but rather to obtain an adapted version equivalent to the "original version". Ideally, there should always be two translators, one without a specific profile focused on semantic equivalence (preferably, a professional translator), and the other with expertise in the content area of the instrument's construct to ensure conceptual equivalence. Translators who are not familiar with the construct of the instrument, more easily identify ambiguous meanings in the original instruments.[26] During the translation of instruments, both translators record their doubts and comments in a form provided by researchers. Before translating, give instructions to translators about the type of translation intended, whether more literal or more cultural. If necessary, also provide materials in the "target language" that facilitate understanding of the construct of the measuring instrument.

Forward Translation Synthesis

In most guidelines for translation, adaptation and cross-cultural validation of instruments, independent translation is articulated with Team-Based Approaches to reach consensus among the involved translators.[91] The method adopted at this stage to synthesize the two translations into a single version is called the Committee Approach.[21,62] This approach consists of a meeting with the translators who participated in the previous step to discuss the translation differences and reach a consensus on the most appropriate translation for each item.[152] The Committee Approach enables the detection of language and culture specific idiosyncrasies from the earliest stages of cross-cultural adaptation of measurement instruments.[153] Some authors[21,46,62] advocate the Committee Approach to be coordinated by a third translator proficient in both languages and without a specific profile. This coordinating element can facilitate the discussion and consensus processes between forward translators. For this purpose, before the meeting takes place, the researchers provide them with a comparative table with both translations and with the doubts and comments issued by each translator. They also ask the third translator to make a translation proposal based on the ambiguities and discrepancies detected in each item to be presented and discussed at the Committee meeting. During the meeting, one of the researchers moderates the discussions and records the decisions made. If it is not possible to involve a third translator, the translation proposal is made by a researcher. The report of this stage should describe the consensual solutions to resolve the ambiguities and discrepancies found. This report must be reviewed by the research team. If necessary, contact the developers of the measuring instrument to clarify possible doubts when translating the items.

Back Translation

Back translation is a quality assurance process to check the accuracy of the forward translation,[91] and theoretically makes it possible to expand unclear wording and "gross inconsistencies/conceptual errors",[21] that need clarification.[62,149] It involves the translation of the instrument from the "target language" (forward translation) into the "original language", resulting in a back translation.[22] Back translators are expected to provide a translation that is as 'literal or faithful as possible',[154] while still respecting the rules of the "target language". They are also required to replicate any mistakes found in the forward translation and note down any discrepancies or non-natural sounding language. These comments are then documented on a form provided by the researchers. As the back translation process is so different from a standard translation process, it is important that translators are trained in what is involved in a back translation. The number and profile of the back translators should be identical to that of the forward translators.[21,62] Almost all guidelines recommend that back translators are familiar with both languages and that they are native speakers of the language of the back translation (which is almost always English). If native speakers are unavailable, there should be at least two translators proficient in the "source language" and in the "target language". More important than including native translators is conducting the back translation in a blinded manner, ie, back translators should not have access to the original measurement instrument,[151] nor being informed about its construct.[21,149] This characteristic ensures conceptual equivalence, however it is not always described in studies. It is particularly important to ensure all back translators are aware of this requirement, as translators are trained to understand the context as much as possible to provide an accurate translation.[150] With the back translation of the measurement instrument, two versions of the instrument in the language

of the original document are generated. Regardless of whether minor discrepancies occur between the two versions, the main aspect that needs to be analyzed in the next step is whether there is a change in meaning between the items in the back translation and the items in the "original instrument".[22]

Harmonization

Similar to the translation synthesis stage, after obtaining the back translated versions, researchers should use a Team-Based Approach. At this stage, all versions of the measurement instrument (the original version, the translated version and the back translated version) are compared by all translators involved in order to identify possible ambiguities and discrepancies, and to decide on the most appropriate translation.[21,62] The Team-Based Approach used at this stage consists of a Multiprofessional Committee.[155] This approach is also referred to as a Committee of Experts,[21] or a Harmonization Meeting.[156] The Multiprofessional Committee consists of a meeting involving members from complementary areas of expertise.[155] To reduce the possibility of content bias resulting from decisions made solely on the basis of semantic equivalence, participants from the target population in which the instrument is to be applied are also included.[151] Sousa et al[62] recommend this Multiprofessional Committee should include at least one member from the research team, one professional familiar with the questionnaire constructs' contents (if possible from the target population), and all the translators involved in translation and back translation, with the exception of the translator who acted as the judge in the synthesis of the translations. It may also involve a monolingual member with mother tongue in the "target language", unfamiliar with the constructs of the instrument to ensure bias reduction. According to Erkut,[157] monolingual members can detect unfamiliar constructions more easily than bilinguals, as they are not influenced by their expertise in the "original language". Contact with the authors of the instruments is also recommended in order to provide their insight into the construction of the instrument and clarify any questions that may arise.[62] Issues to be clarified with authors may result from a disagreement on the translation of certain items.[91] Some authors suggest the inclusion of a linguistic expert to ensure idiomatic and semantic equivalence.[62] Before the Multiprofessional Committee meets, one of the researchers compares the back translations with the translations and with the "original instrument" in order to identify ambiguities and discrepancies, which will be presented and discussed in the Committee.[22] If a linguistic expert participates, this activity may be requested from this expert, allowing the researcher to focus on the discussions during the meeting and the documentation of the agreed-upon solutions.

The Multiprofessional Committee aims to obtain consensus among all experts regarding: 1) possible ambiguities and discrepancies related to cultural meanings; 2) colloquialisms; 3) phrases and words and 4) idiomatic expressions.[62] Regarding idiomatic expressions, consensus may not be easy. These are combinations of words which may not be easy to translate,[158] because they have a specific cultural meaning different from their semantic meaning.[159] One of the strategies that can be followed in the harmonization of idiomatic expressions, is to identify comparable idiomatic expressions within the "target culture".[160] Participants from the target population can identify idiomatic expressions that are used within their culture easier than translators with no specific profile. Inputs from participants of the target population are very important to ensure conceptual and functional equivalence of items where ambiguities and discrepancies were found.[151] The decision procedure on the harmonization of items that seeks to articulate the search for conceptual equivalence with functional equivalence is the Dual-Focus.[75,157,161] Its use is important in cross-cultural validation studies where the instruments use a specific terminology with which some translators are not familiar. Several studies reporting the substitution of words and items as a consequence of the cross-cultural adaptation process have underlined this procedure.[77–83] For partial or total replacement of items, researchers must obtain approval from the developers of the "original instrument".

Pre-testing

A rigorous back translation may be insufficient to guarantee that all semantic and conceptual discrepancies are resolved. Carrying out a pre-testing provides the identification of problems that may affect the reliability and validity of the translated version of the instrument,[162] namely related to the clarity and relevance of the instrument's items.[163] In the case of attitude measurement instruments, pre-testing is important as questions about attitudes can be sensitive in the context to which they refer.[39] In this step, researchers assess the suitability of the measurement instrument before its using in the field test.[164] A pre-test involves data collection from a small number of participants of the target population,[163,165] typically using the same sampling method as planned for the study or, alternatively, the method of

convenience sampling.[165] Anyhow, the inclusion of members of the target population may involve a risk of contamination of the study sample, since participants who have been exposed to the pre-test may respond differently from those who have not had this experience,[164] which can be avoided by excluding these elements from the sample. When pre-testing measurement instruments, two complementary pre-tests are ideally recommended.[21,62,166]

The first pre-test involves using the Interview with Cognitive Debriefing to assess the clarity of the measurement instrument, particularly of the items.[62] The Interview with Cognitive Debriefing comprises a set of questions that are addressed individually to a set of participants after they have answered the instrument.[25,167] In the instrument, researchers ask participants to rate the clarity of all items using a dichotomous scale ("it is clear"; "it is not clear").[62] Alternatively, could be used a trichotomous rating scale ("it is not clear"; "item needs some revision", and"very clear").[168] After answering, researchers ask improvement suggestions to improve the wording in relation to the items marked as "unclear" or 'needs some revision'. Because it involves a recommended sample size of 30 participants,[169] it is a highly time-consuming pre-test.[170] The second pre-test uses an Expert Panel to critically evaluate the measurement instrument,[170,171] regarding the items relevance.[62] The relevance of the items is assessed with a 4-point Likert scale (1. "not at all relevant", 2. "somewhat relevant", 3. "quite relevant" and 4. "highly relevant",[172,173] or with a 3-point Likert scale ('not at all relevant', 'somewhat relevant', and 'very relevant').[174] Once a high number of experts may reduce the possibility of concordance,[175] we propose to adopt the Almanasreh's recommendation of 5 to 10 experts.[176] The Experts Panel is a procedure specially suitable in cultural adaptation of instruments that uses a highly specific terminology.[177] Alternatively to the Experts Panel, the second pre-test may involve to conduct a Focus Group.[25] This is a technique usually preceded by a questionnaire to prepare a discussion and to provide additional data for subsequent analysis.[178] The Focus Group moderator must be able to balance the contributions of all participants to keep the discussion going and interpret the information correctly. This means he/she must avoid consensus biased by the group dynamics.[170,178] Regardless the methodological approach, both pre-tests may involve qualitative and/or quantitative methods.[164] The diversity of methods in pre-testing improves the quality of cultural adaptation of instruments.[170,179]

For both the assessment of items clarity and the assessment of its relevance, a minimum level of agreement among participants is required for each item. Based on the results of each pre-test, researchers may consider revising these items accordingly.[25] This requires researchers to check whether translation doubts were reported for those particular item(s) and how these doubts were resolved.[21] Consulting the documentation of the previous steps makes it possible to exclude a semantic equivalence problem in order to replace or eliminate items from the measurement instrument.[21] To support the decision, the data obtained during the pre-test stage can be submitted to a statistical analysis regarding the consistency and accuracy of the degree of agreement between reviewers.[163] This analysis can be performed by calculating the CVI to quantify the content validity of the adapted version of the instrument.[180] It is suitable for dichotomous answers but can also be used for Likert-type multiple-choice response formats by recoding the answers.[181] To calculate the CVI, researchers may use two approaches: 1) calculating each item's content validity index (I-CVI) and 2) calculating the mean of the CVI of all items included in the instrument (S-CVI/Ave).[173] According to Polit et al,[180] items with a I-CVI near of 0.78 must be revised and items with low I-CVI must be excluded. Polit et al[182] recommend that for an instrument to be judged as having excellent content validity, the items should have I-CVIs ≥0.78 and S-CVI/Ave ≥0.90. We recommend that all items with a CVI< 0.78 be reviewed and reevaluated by the research team and members of the target population.

Content Validity Ratio (CVR) is another method to quantify the content validity of dichotomous ratings on items. This is an approach proposed by Lawshe,[183] which includes a critical number of experts rating the relevance of individual items as "essential", "useful but not essential" or "not necessary", and those items considered "essential" are included in the instrument.[176] CVR values range from −1 (perfect disagreement) to 5 (perfect agreement). Values above zero indicate that more than 50% of the panel experts agree the item is essential.[176] When interpreting the CVR, it is necessary to consider whether the level of agreement among the experts is above what may have occurred by chance.[184] In this regard, the critical CVR values presented by Ayre & Scally,[184] can be considered by novice researchers to determine how many panel experts need to agree an item is essential and decide which items will be included or reviewed based on CVR values. For example, considering a panel of 10 experts, at least nine must agree the item is "essential", a critical CVR of 0.8 should be considered, items with CVR ≥ 0.8 would be included and those with CVR < 0.8 be reviewed and reevaluated..

For categorial scales the Kappa Coefficient of Concordance (K) must be estimated.[62] This coefficient allows a more independent assessment once it expresses a degree of inter-rater agreement devoid of the proportion of agreement that results from chance.[185] Therefore, it is an important supplement to CVI values.[172] Its use is suitable only for dichotomous data.[172] Kappa Coefficient is calculated from the following formula: K=[(a+b)(a+c)]+[(c+d)(b+d)] and varies theoretically between −1 e 1.[185] A K between 0.60 and 0.74 is an indicator of a good level of agreement and between 0.75 and 1.0 indicates an excellent level.[186,187] The minimum acceptable value of K is a coefficient of agreement of 0.60.[173] The interpretation of the inter-rater reliability of the instrument may result from the CVI and Kappa Coefficient values, together with the calculation of the ICC obtained later in the psychometric validation stage.[173] Consequently, the decision of deleting items can be postponed to this stage by the results of the ICC. Despite this, the CVI and Kappa results can be interpreted in terms of the underlying factors and the measures that can be taken to improve them.[188] Regardless of whether the decision is deferred, whenever there is a need to revise reevaluate items, with partial or total substitutions and eliminations, researchers should ensure that such procedures do not compromise the construct coverage of the original instrument.[163]

Field Testing

This step involves preparing the pre-final version of the measurement instrument for data collection in the target population and the actual data collection. The preparation of the instrument may include some decisions, namely: 1) reversing the response formats of items that were negatively phrased, and 2) calculating a minimum sample size that enables the psychometric validation. Negatively worded item responses need to be reversed in order to not affect the evaluation of the reliability coefficient with negative inter-item correlations.[189] To determine the minimum sample size, researchers can use the criterion required for conducting factor analysis of five to 10 subjects per instrument item.[190] The higher this ratio, the greater the possibility of obtaining a robust factor structure model.

Psychometric Validation

Psychometric validation is the metric or empirical validation of a measurement instrument. To be considered valid, an instrument needs to produce the same results under the same conditions.[191] This validation comprises: 1) performing an EFA; 2) analyzing the internal consistency and 3) performing a CFA. These procedures require researchers to have knowledge of statistical data processing and to use specific software for this purpose. The first analysis to be performed is the EFA, which groups the items of the measurement instrument into factors or dimensions based on their correlations. Conducting a literature review or concept analysis study can help researchers make an informed decision about the number of factors obtained through EFA. Then, these factors can be processed as new meaningful variables and are theoretically named by researchers.[192] EFA is recognized as an abductive method of theory generation, which is further evaluated by CFA.[193,194] The abductive method is characterized by the use of analogy to construct descriptions and theoretical explanations of reality.[195] It also allows researchers to conceptually refine the phenomenon under study after validating a particular measurement instrument. The exclusion of items from the "original instrument" provides an opportunity for researchers to refine their conceptualization about the phenomenon they are studying.[196] Despite that, the decision to retain or exclude items rests with the researcher, supported by the contextual analysis of the instrument´s construct.

Once obtained the factor structure of the measurement instrument, researchers analyze its internal consistency, which indicates the degree of repeatability of its results.[197] The most common method for assessing internal consistency is through the Cronbach's α, which measures the degree of correlation between items.[198] The interpretation from the Cronbach's α can be based on the following ranges: 1) $\alpha<0.60$ (weak value); 2) $0.60\leq\alpha<0.70$ (questionable value); 3) $0.70\leq\alpha<0.80$ (acceptable value); 4) $0.80\leq\alpha<0.90$ (good value) and 5) $\alpha\geq0.90$ (excellent value).[199] Some authors consider acceptable a Cronbach's α of 0.70 for measurement instrument that are being refined and tested in culturally different samples.[200] Others consider that in Social Sciences it is acceptable a Cronbach's α ranging between 0.60 and 0.70 in exploratory studies.[201] Despite being the most widely used index for assessing internal consistency, Cronbach's α tends to underestimate the total reliability of a measure, estimating reliability conservatively. One approach to address this issue is to promote homogeneity by standardizing items before calculating the index or to work directly with correlation coefficients (standardized covariance), which results in a standardized Cronbach's α index. To refine the analysis, it can also be calculated the inter-item correlation coefficients and

the item-total correlation.[202] Values between 0.15 and 0.50 for the inter-item coefficients are considered acceptable for comprehensive constructs.[203] For the item-total coefficients, values above 0.30 are considered acceptable.[204] When the factor structure is multidimensional, researchers analyze the internal consistency in relation to each of the factors and in relation to the total instrument. The results obtained are interpreted and compared with the values reported by the authors of the "original version".

CFA allows researchers to specify how the final structure of the measurement instrument should look like.[205] This statistical technique starts from a hypothesized factor structure obtained in the EFA and from the participants' data to analyze the feasibility of this structure.[206] The quality of the global adjustment of the factor structure model can be assessed based in Marôco's values of reference,[207] regarding: 1) Chi-square test (X^2/df, whose reference value should be as low as possible); 2) Comparative Fit Index (CFI) where values ≥ 0.90 and ≥ 0.95 reflect, respectively, a good and very good fit; 3) Goodness of Fit Index (GFI), whose reference values are the same as the CFI; 4) Root Mean Square Error of Approximation (RMSEA), which must be >0.10; 5) significance level of RMSEA, P[rmsea]≤0.005 and 6) Modified Expected Cross-Validation Index (MECVI), being desirable the lowest possible value. The confirmation of the adequacy of the factor model of a measurement instrument that has excluded items from the "source version" may lead researchers to propose a new model or conceptual framework for the phenomenon under study. The conceptual model has a limited scope of explaining a phenomenon or part of it, whereas a conceptual framework represents the phenomenon in a descriptive network of interconnected concepts that eases its understanding.[208]

Psychometric Properties Analysis

The assessment of the adequacy of any assessment instrument requires the analysis of its purpose, conceptual basis, development, and psychometric properties. The psychometric properties of measurement instruments correspond to a set of evidences produced by researchers during the process of translation, adaptation and cross-cultural validation and psychometric, which allow to assess the validity and reliability of the results obtained.[209] There are different classifications in the literature for the psychometric properties of measurement instruments.[8,210] The Consensus Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) methodological quality assessment descriptors provide a structured guide to the evidence that enables researchers to analyze the psychometric properties of the measurement instrument at the end of their study.[136] This analysis should focus on at least three psychometric properties: 1) content validity; 2) construct validity and 3) internal consistency. Content validity analyses the degree to which an instrument reflects the domain of interest and the conceptual definition of a construct.[101] Within content validity, face validity can be assessed, which allows understanding, through the opinions of the reviewers who participated in the pretests, whether the instrument actually assesses what its authors claim it does.[211–213] Despite the quantification of the agreement rates, face validity should not be considered alone as a factual indicator of validity of measurement instruments.[212] Construct validity refers to the degree to which measurement instruments enable to produce legitimate inferences to be drawn from scores for theoretical constructs and on which these observations are supported.[214] This is a central feature in the process of validating measurement instruments and encompasses important sources of evidence: 1) the evidence based on the content of the test, and 2) the evidence based on the internal structure.[215] The inter-item and total-item coefficients and the degree of adjustment of the factor model allow judgments to be made about construct validity. Cross-cultural validity is also an element of construct validity.[216] It indicates the extent to which the performance of items in a culturally adapted instrument reflects that in the "original version".[136] The performance of independent, blinded translations and the use of Team-Based Approaches are examples that contribute to this validity. Depending on the study design, researchers may plan to evaluate other psychometric properties that were not reported by the developers of the "original instrument".

Implications for Practice

The process of translation, adaptation, and cross-cultural validation is a long and complex process during which researchers may be unaware of the possible methodological approaches available to overcome possible barriers. This study provides comprehensive information on the methodological approaches recommended by existing guidelines in the field of healthcare sciences, to support novice researchers in planning cross-cultural validation studies of measurement instruments. In addition, it

allows these researchers to focus on the knowledge and skills about qualitative and quantitative methodological approaches that they need to acquire and develop to be able to conduct a validation study. Finally, it allows novice researchers to develop team leadership skills in the research process. This includes not only coordinating the activities of the team of translators and experts that comprise the research team but also foreseeing the research outputs can be generated with the process of translation, adaptation, and cross-cultural validation. Examples of such outputs include a literature review article on the construct of the instrument, a concept analysis article, a methodological article reporting the process of translation and cross-cultural adaptation, and another detailing the psychometric properties' quality of the instrument.

Conclusion

This is the first methodological literature review conducted in the healthcare sciences that focuses on the methodological approaches recommended by existing guidelines. Based on this review, a practical guideline was developed to assist researchers in the decision-making process during validation studies. It is a practical guideline because it includes extensive and comprehensive information about alternative methodological approaches that researchers can use throughout the process. This information enables researchers to manage the complexity of cross-cultural validation studies and obtain an instrument with recognized psychometric qualities. Although we have limited the literature review to the field of healthcare sciences, the guideline provided also applies to the translation, adaptation, and validation of measurement instruments in other scientific fields. In the future, we recommend researching the impacts of using our practical guideline on researchers' experience and outcomes when conducting cross-cultural validation studies.

Acknowledgments

We would like to thank the Handovers4SafeCare® project of the CIDNUR for leading the development of this study.

Disclosure

The authors report no conflicts of interest in this work.

References

1. May C, Finch T, Rapley T. Normalization process theory. In: Nielsen P, Birken SA, editors. *Handbook of Implementation Science*. Edward Elgar Publishing; 2020:157.
2. Portney LG. Designing surveys and questionnaires. In: *Foundations of Clinical Research: Applications to Evidence-Based Practice*. 4th ed. F.A. Davis Company; 2020:141–150.
3. Chow S-C, Liu J-P. Basic statistic concepts. In: *Design and Analysis of Clinical Trials*. 2nd ed. Wiley-Interscience; 2004:47–54.
4. Sidani S, Braden CJ. Examination of interventions´ acceptance. In: *Nursing and Health Interventions: Design, Evaluation, and Implementation*. 2nd ed. Wiley-Blackwell; 2021:218–225.
5. Efstathiou G. Translation, adaptation and validation process of research intruments. In: Suohnen R, Stolt M, Papatravou E, editors. *Individualized Care: Theory, Measurement, Research and Practice*. Springer; 2019.
6. Cha E-S, Kim KH, Erlen JA. Translation of scales in cross-cultural research: issues and techniques. *J Adv Nurs*. 2007;58(4):386–395. doi:10.1111/j.1365-2648.2007.04242.x
7. Sutherland S. Outcomes research. In: Gray JR, Grove SK, editors. *The Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence*. 9th ed. Elsevier; 2020:365.
8. Miller LA, Lovler RL. How do we assess the psychometric quality of a test? In: *Foundation of Psychological Testing: A Practical Approach*. 6th ed. Sage Publications; 2020:326.
9. Kristjansson EA, Desrochers A, Zumbo B. Translating and adapting measurement instruments for cross-linguistic and cross-cultural research: a guide for practitioners. *Can J Nurs Res*. 2003;35(2):127–142.
10. Alagappan T. The cross-cultural adaptation process of a patient-reported outcome measure. *J Sci Soc*. 2023;50(1):13. doi:10.4103/jss.jss_136_21
11. Sul SIR, Lucas PRMB. Translation and validation of the anticipated turnover scale for the Portuguese cultural context. *Nurs Open*. 2020;7(5):1475–1481. doi:10.1002/nop2.521
12. Almeida S, Nascimento A, Lucas PB, Jesus É, Araújo B. RN4CAST study in Portugal: validation of the Portuguese version of the Practice Environment Scale of the Nursing Work Index. *Aquichan*. 2020;20(3):1–10. doi:10.5294/aqui.2020.20.3.8
13. Lucas P, Jesus E, Almeida S, Araújo B. Validation of the psychometric properties of the Practice Environment Scale of Nursing Work Index in primary health care in Portugal. *Int J Environ Res Public Health*. 2021;18(12):6422. doi:10.3390/ijerph18126422
14. Anunciada S, Benito P, Gaspar F, Lucas P. Validation of psychometric properties of the Nursing Work Index: revised scale in Portugal. *Int J Environ Res Public Health*. 2022;19(9):4933. doi:10.3390/ijerph19094933
15. Carvalho M, Gaspar F, Potra T, Lucas P. Translation, adaptation, and validation of the Self-efficacy scale for clinical nurse leaders for the Portuguese culture. *Int J Environ Res Public Health*. 2022;19(14):8590. doi:10.3390/ijerph19148590

16. Sousa E, Lin C-F, Gaspar F, Lucas P. Translation and validation of the indicators of quality nursing work environments in the Portuguese cultural context. *Int J Environ Res Public Health*. 2022;19(19):12313. doi:10.3390/ijerph191912313
17. Cabrita C, Lucas P, Teixeira G, Gaspar F. Translation and validation of the Individual Workload Perception Scale: revised for Portuguese nurses. *Healthcare*. 2022;10(12):2476. doi:10.3390/healthcare10122476
18. Gomes P, Ribeiro S, Silva M, et al. Cross-cultural validation of the Portuguese version of the Quality of Oncology Nursing Care Scale. *Cancers (Basel)*. 2024;16(5):859. doi:10.3390/cancers16050859
19. Cunha F, Pinto MR, Riesch S, Lucas P, Almeida S, Vieira M. Translation, adaptation, and validation of the Portuguese version of the Exercise of Self-Care Agency Scale. *Healthcare*. 2024;12(2):159. doi:10.3390/healthcare12020159
20. Waltz CF, Strickland OL, Lenz ER. Other mesurement issues. In: *Measurement in Nursing and Health Research*. 4th ed. Springer Publishing; 2010:446–448.
21. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–3191. doi:10.1097/00007632-200012150-00014
22. Hedrih V. Test translation. In: *Adapting Psychological Tests and Measurement Instruments for Cross-Cultural Research: An Introduction*. Routledge; 2020:48–98.
23. Nord C. Defining translation functions: the translation brief as a guideline for the trainee translator. *Ilha Do Desterro*. 1997;33:41–55.
24. Montalt V, González-Davies M. Understanding medical communication. In: *Medical Translation Step by Step: Learning by Drafting*. Routledge; 2006:47.
25. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*. 2015;68(4):435–441. doi:10.1016/j.jclinepi.2014.11.021
26. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*. 1998;7(4):323–335. doi:10.1023/a:1024985930536
27. Peña ED. Lost in translation: methodological considerations in cross-cultural research. *Child Dev*. 2007;78(4):1255–1264. doi:10.1111/j.1467-8624.2007.01064.x
28. Alonso J, Black C, Norregaard J-C, et al. Cross-cultural differences in the reporting of global functional capacity. *Med Care*. 1998;36 (6):868–878. doi:10.1097/00005650-199806000-00010
29. Muniz J, Hambleton RK, Xing D. Small samples studies to detect flaws in item translations. *Int J Test*. 2001;1(2):115–135. doi:10.1207/S15327574IJT0102_2
30. Waltz CF, Strickland OL, Lenz ER. Other measurement issues. In: *Measurement in Nursing and Health Research*. 4th ed. Springer Publishing; 2010:449–452.
31. Riccio CA, Yoon H, McCormick AS. Neuropsychological test selection with clients who are Asian. In: Davis JM, D´Amato RC, editors. *Neuropsychology of Asians and Asian-Americans: Practical and Theoretical Considerations*. Spinger Publishing; 2014:153.
32. Smith TW, et al. Developing and evaluating cross national survey instruments. In: Presser S, Rothgeb JM, Couper MP, editors. *Methods for Testing and Evaluating Survey Questionnaires*. Wiley Interscience; 2004. 439–442.
33. van de Vijver F, Portinga YH. Conceptual and methodological issues in adapting tests. In: Hembletom RK, Merenda PF, Spilberg CD, editors. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates; 2005:41–47.
34. Austin PC, Brunner LJ. Type I error inflation in the presence of a ceiling effect. *Am Stat*. 2003;57(2):97–104. doi:10.1198/0003130031450
35. Holbrook A. Acquiescence response bias. In: Lavrakas PJ, editor. *Encyclopedia of Survey Research Methods*. Vol. I. Sage; 2008:3.
36. van de Vijver FJR, Poortinga YH. Towards an Integrated Analysis of Bias in Cross-Cultural Assessment. *Eur J Psychol Assess*. 1997;13 (1):29–37. doi:10.1027/1015-5759.13.1.29
37. Bartram D. Increasing validity with forced-choice criterion measurement formats. *Int J Sel Assess*. 2007;15(3):263–272. doi:10.1111/j.1468-2389.2007.00386.x
38. Wivagg J. Forced choice. In: Lavrakas PJ, editor. *Encyclopedia of Survey Research Methods*. Sage; 2008:289–290.
39. Maitland A. Attitudes measurement. In: Lavrakas PJ, editor. *Encyclopedia of Survey Research Methods*. Sage; 2008:37–38.
40. Hilton A, Skrutkowski M. Translating instruments into other languages: development and testing processes. *Cancer Nurs*. 2002;25(1):1–7. doi:10.1097/00002820-200202000-00001
41. Lino de CR, Brüggemann M, Souza M de L de OM, Barbosa de S, Santos Dos EKA. Adaptação transcultural de instrumentos de pesquisa conduzida pela enfermagem do Brasil: uma revisão integrativa. *Texto Context - Enferm*. 2018;26(4). doi:10.1590/0104-07072017001730017
42. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417–1432. doi:10.1016/0895-4356(93)90142-N
43. Machado da R, Fernandes AD de BF S, Oliveira de ALCB, Soares LS, Gouveia de O MT, Silva da GRF. Métodos de adaptação transcultural de instrumentos na área da enfermagem. *Rev Gaúcha Enferm*. 2018;39. doi:10.1590/1983-1447.2018.2017-0164
44. Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate Health-Related Quality of Life Questionnaires for use in multinational clinical trials. *Value Heal*. 2008;11(3):509–521. doi:10.1111/j.1524-4733.2007.00292.x
45. Farina N, Jacobs R, Sani TP, et al. Description of the cross-cultural process adopted in the STRiDE (STrengthening Responses to dementia in DEveloping countries) program: a methodological overview. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2022;14(1). doi:10.1002/dad2.12293
46. Furukawa R, Driessnack M, Colclough Y. A committee approach maintaining cultural originality in translation. *Appl Nurs Res*. 2014;27 (2):144–146. doi:10.1016/j.apnr.2013.11.011
47. Helmich E, Cristancho S, Diachun L, Lingard L. 'How would you call this in English?': being reflective about translations in international, cross-cultural qualitative research. *Perspect Med Educ*. 2017;6(2):127–132. doi:10.1007/S40037-017-0329-1
48. Uysal-Bozkir Ö, Parlevliet JL, de Rooij SE. Insufficient cross-cultural adaptations and psychometric properties for many translated health assessment scales: a systematic review. *J Clin Epidemiol*. 2013;66(6):608–618. doi:10.1016/j.jclinepi.2012.12.004
49. Albach CA, Wagland R, Hunt KJ. Cross-cultural adaptation and measurement properties of generic and cancer-related patient-reported outcome measures (PROMs) for use with cancer patients in Brazil: a systematic review. *Qual Life Res*. 2018;27(4):857–870. doi:10.1007/s11136-017-1703-5
50. Praveen S, Parmar J, Chandio N, Arora A. A systematic review of cross-cultural adaptation and psychometric properties of oral health literacy tools. *Int J Environ Res Public Health*. 2021;18(19):10422. doi:10.3390/ijerph181910422

51. Min SN, Duangthip D, Gao SS, Detsomboonrat P. Quality of the adaptation procedures and psychometric properties of the scale of oral health outcomes for 5-year-old children (SOHO-5): a systematic review. *Qual Life Res*. 2023;32(6):1537–1547. doi:10.1007/s11136-022-03280-2
52. Danielsen AK, Pommergaard H-C, Burcharth J, Angenete E, Rosenberg J. Translation of questionnaires measuring health related quality of life is not standardized: a literature based research study. *PLoS One*. 2015;10(5):e0127050. doi:10.1371/journal.pone.0127050
53. Øygarden A-MU, Berg RC, Abudayya A, Glavin K, Strøm BS. Measurement instruments for parental stress in the postpartum period: a scoping review. *PLoS One*. 2022;17(3):e0265616. doi:10.1371/journal.pone.0265616
54. Echevarría-Guanilo ME, Gonçalves N, Romanoski PJ. Psychometric properties of measurement instruments: conceptual basis and evaluation methods: part II. *Texto Context - Enferm*. 2019;28(e20170311):1–14. doi:10.1590/1980-265x-tce-2017-0311
55. Cruchinho P, Teixeira G, Lucas P, Gaspar F. Evaluating the methodological approaches of cross-cultural adaptation of the Bedside Handover Attitudes and Behaviours Questionnaire into Portuguese. *J Healthc Leadersh*. 2023;15:193–208. doi:10.2147/JHL.S422122
56. Slade D, Murray KA, Pun JKH, Eggins S. Nurses' perceptions of mandatory bedside clinical handovers: an Australian hospital study. *J Nurs Manag*. 2019;27(1):161–171. doi:10.1111/jonm.12661
57. Freudental-Pederson M, Hartmann-Peterson K, Nielsen LD. Mixing methods in the search for mobile complexity. In: Ficham B, Mark M, Murray L, editors. *Mobile Methodologies*. Palgrave McMillan; 2010:30.
58. Peters M, Passchier J. Translating instruments for cross-cultural studies in headache research. *Headache J Head Face Pain*. 2006;46(1):82–91. doi:10.1111/j.1526-4610.2006.00298.x
59. Weeks A, Swerissen H, Belfrage J. Issues, challenges, and solutions in translating study instruments. *Eval Rev*. 2007;31(2):153–165. doi:10.1177/0193841X06294184
60. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Heal*. 2005;8(2):94–104. doi:10.1111/j.1524-4733.2005.04054.x
61. Reichenheim ME, Moraes CL. Operacionalização de adaptação transcultural de instrumentos de aferição usados em epidemiologia. *Rev Saude Publica*. 2007;41(4):665–673. doi:10.1590/S0034-89102006005000035
62. Sousa V, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract*. 2011;17(2):268–274. doi:10.1111/j.1365-2753.2010.01434.x
63. Hagell P, Hedin P-J, Meads DM, Nyberg L, McKenna SP. Effects of Method of Translation of Patient-Reported Health Outcome Questionnaires: a Randomized Study of the Translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. *Value Heal*. 2010;13(4):424–430. doi:10.1111/j.1524-4733.2009.00677.x
64. Lee WL, Chinna K, Lim Abdullah K, Zainal Abidin I. The forward-backward and dual-panel translation methods are comparable in producing semantic equivalent versions of a heart quality of life questionnaire. *Int J Nurs Pract*. 2019;25(1). doi:10.1111/ijn.12715
65. Papadakis NM, Aletta F, Kang J, Oberman T, Mitchell A, Stavroulakis GE. Translation and cross-cultural adaptation methodology for soundscape attributes – a study with independent translation groups from English to *Greek*. *Appl Acoust*. 2022;200:109031. doi:10.1016/j.apacoust.2022.109031
66. van de Vijver FJR, Poortinga YH. On the study of culture in developmental science. *Hum Dev*. 2002;45(4):246–256. doi:10.1159/000064985
67. Petkovic J, Epstein J, Buchbinder R, et al. Toward ensuring health equity: readability and cultural equivalence of OMERACT patient-reported outcome measures. *J Rheumatol*. 2015;42(12):2448–2459. doi:10.3899/jrheum.141168
68. Bundgaard K, Nisbeth Brøgger M. "Don't fix bad translations": a netnographic study of translators' understandings of back translation in the medical domain. *MonTi Monogr Traducción e Interpret*. 2018;1(10):205–224. doi:10.6035/MonTI.2018.10.8
69. Hawkins M, Cheng C, Elsworth GR, Osborne RH. Translation method is validity evidence for construct equivalence: analysis of secondary data routinely collected during translations of the Health Literacy Questionnaire (HLQ). *BMC Med Res Methodol*. 2020;20(1):130. doi:10.1186/s12874-020-00962-8
70. Epstein J, Osborne RH, Elsworth GR, Beaton DE, Guillemin F. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. *J Clin Epidemiol*. 2015;68(4):360–369. doi:10.1016/j.jclinepi.2013.07.013
71. Teig CJP, Bond MJ, Grotle M, et al. A novel method for the translation and cross-cultural adaptation of health-related quality of life patient-reported outcome measurements. *Health Qual Life Outcomes*. 2023;21(1):13. doi:10.1186/s12955-023-02089-y
72. Tsai T-I, Luck L, Jefferies D, Wilkes L. Challenges in adapting a survey: ensuring cross-cultural equivalence. *Nurse Res*. 2018;26(1):28–32. doi:10.7748/nr.2018.e1581
73. Jayawickreme E, Jayawickreme N, Goonasekera MA. Using focus group methodology to adapt measurement scales and explore questions of wellbeing and mental health. *Intervention*. 2012;10(2):156–167. doi:10.1097/WTF.0b013e328356f3c4
74. Montenegro M, Valdez D, Crawford B, Turner R, Lo W-J, Jozkowski KN. Using a decentering framework to create English/Spanish surveys about abortion: insights into comparative survey research (CSR) for new survey development and recommendations for optimal use. *Soc Sci J*. 2022;1–15. doi:10.1080/03623319.2022.2092379
75. Vizcaya-Moreno MF, Pérez-Cañaveras RM. Country validation of the CLES-Scale: linguistic and cultural perspectives. In: Saarikoski M, Strandell-Laine C, editors. *The CLES-Scale: An Evaluation Tool for Healthcare Education*. Springer; 2018:31–46.
76. Erkut S. Developing multiple language versions of instruments for intercultural research. *Child Dev Perspect*. 2010;4(1):19–24. doi:10.1111/j.1750-8606.2009.00111.x
77. Marcondes FB, de Vasconcelos RA, Marchetto A, de Andrade ALL, Filho AZ, Etchebehere M. Translation and cross-cultural adaptation of the Rowe score for Portuguese. *Acta Ortop Bras*. 2012;20(6):346–350. doi:10.1590/S1413-78522012000600007
78. Nepal GM, Shrestha A, Acharya R. Translation and cross-cultural adaptation of the Nepali version of the Rowland Universal Dementia Assessment Scale (RUDAS). *J Patient-Reported Outcomes*. 2019;3(1). doi:10.1186/s41687-019-0132-3
79. Coll-Risco I, Camiletti-Moirón D, Acosta-Manzano P, Aparicio VA. Translation and cross-cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) into Spanish. *J Matern Neonatal Med*. 2019;32(23):3954–3961. doi:10.1080/14767058.2018.1479849
80. Pereira GIDN, Costa CDDS, Geocze L, Borim AA, Ciconelli RM, Camacho-Lobato L. Cross-cultural adaptation and validation for Portuguese (Brazil) of health related quality of life instruments specific for gastroesophageal reflux disease. *Arq Gastroenterol*. 2007;44(2):168–177. doi:10.1590/s0004-28032007000200016

81. Grundström H, Rauden A, Olovsson M. Cross-cultural adaptation of the Swedish version of Endometriosis Health Profile-30. *J Obstet Gynaecol (Lahore)*. 2020;40(7):969–973. doi:10.1080/01443615.2019.1676215
82. Polesello GC, Godoy GF, De Castro Trindade CA, De Queiroz MC, Honda E, Ono NK. Translation and cross-cultural adaptation of the modified Hip outcome tool (mhot) into Portuguese. *Acta Ortop Bras*. 2012;20(2):88–92. doi:10.1590/S1413-78522012000200006
83. Muquith MA, Islam MN, Haq SA, Ten Klooster PM, Rasker JJ, Yunus MB. Cross-cultural adaptation and validation of a Bengali version of the modified fibromyalgia impact questionnaire. *BMC Musculoskelet Disord*. 2012;13(1):1. doi:10.1186/1471-2474-13-157
84. Toma G, Guetterman TC, Yaqub T, Talaat N, Fetters MD. A systematic approach for accurate translation of instruments: experience with translating the Connor–Davidson Resilience Scale into Arabic. *Methodol Innov*. 2017;10(3):205979911774140. doi:10.1177/2059799117741406
85. Poot CC, Meijer E, Fokkema M, Chavannes NH, Osborne RH, Kayser L. Translation, cultural adaptation and validity assessment of the Dutch version of the eHealth Literacy Questionnaire: a mixed-method approach. *BMC Public Health*. 2023;23(1):1006. doi:10.1186/s12889-023-15869-4
86. Hasani L, Santoso H, Junus K. Instrument development for investigating students' intention to participate in online discussion forums: cross-cultural and context adaptation sing SEM. *J Educ Online*. 2021;18(3). doi:10.9743/JEO.2021.18.3.9
87. Bundgaard K, Brøgger MN. Who is the back translator? An integrative literature review of back translator descriptions in cross-cultural adaptation of research instruments. *Perspectives (Montclair)*. 2019;27(6):833–845. doi:10.1080/0907676X.2018.1544649
88. Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess*. 1994;6(4):304–312. doi:10.1037/1040-3590.6.4.304
89. van Widenfelt BM, Treffers PDA, de Beurs E, Siebelink BM, Koudijs E. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev*. 2005;8(2):135–147. doi:10.1007/s10567-005-4752-1
90. Son J. Back translation as a documentation tool. *Int J Transl Interpret Res*. 2018;10(2):89–100. doi:10.12807/ti.110202.2018.a07
91. Coster WJ, Mancini MC. Recommendations for translation and cross-cultural adaptation of instruments for occupational therapy research and practice. *Rev Ter Ocup da Univ São Paulo*. 2015;26(1):50. doi:10.11606/issn.2238-6149.v26i1p50-57
92. Ortiz-Gutiérrez S, Cruz-Avelar A. Translation and cross-cultural adaptation of health assessment tools. *Actas Dermo-Sifiliográficas (English Ed*. 2018;109(3):202–206. doi:10.1016/j.adengl.2018.02.003
93. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud Publica Mex*. 2013;55(1):57–66. doi:10.1590/S0036-36342013000100009
94. Prakash V, Shah S, Hariohm K. Cross-cultural adaptation of patient-reported outcome measures: a solution or a problem? *Ann Phys Rehabil Med*. 2019;62(3):174–177. doi:10.1016/j.rehab.2019.01.006
95. Fortes CPDD, Araújo de QC AP. Check list para tradução e adaptação transcultural de questionários em saúde. *Cad Saúde Coletiva*. 2019;27 (2):202–209. doi:10.1590/1414-462x201900020002
96. Hernández A, Hidalgo MD, Hambleton RK, Gómez-Benito J. International test commission guidelines for test adaptation: a criterion checklist. *Psicothema*. 2020;32(3):390–398. doi:10.7334/psicothema2019.306
97. DuBay M, Watson LR. Translation and cultural adaptation of parent-report developmental assessments: improving rigor in methodology. *Res Autism Spectr Disord*. 2019;62:55–65. doi:10.1016/j.rasd.2019.02.005
98. Hall DA, Zaragoza Domingo S, Hamdache LZ, et al. A good practice guide for translating and adapting hearing-related questionnaires for different languages and cultures. *Int J Audiol*. 2018;57(3):161–175. doi:10.1080/14992027.2017.1393565
99. José M, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*. 2013;25(2):151–157. doi:10.7334/psicothema2013.24
100. Santo RM, Ribeiro-Ferreira F, Alves MR, Epstein J, Novaes P. Enhancing the cross-cultural adaptation and validation process: linguistic and psychometric testing of the Brazilian–Portuguese version of a self-report measure for dry eye. *J Clin Epidemiol*. 2015;68(4):370–378. doi:10.1016/j.jclinepi.2014.07.009
101. Arafat S, Chowdhury H, Qusar M, Hafez M. Cross cultural adaptation and psychometric validation of research instruments: a methodological review. *J Behav Heal*. 2016;5(3):129. doi:10.5455/jbh.20160615121755
102. Kyriazos TA. Applied psychometrics: the 3-Faced Construct Validation Method, a routine for evaluating a factor structure. *Psychology*. 2018;09 (08):2044–2072. doi:10.4236/psych.2018.98117
103. Plonsky L, Kim Y. Task-based learner production: a substantive and methodological review. *Annu Rev Appl Linguist*. 2016;36:73–97. doi:10.1017/S0267190516000015
104. Peters MD, Godfrey C, McInerney P, Soares CB, Khalil H, Parker D. Development of a scoping review protocol. In: Aromataris E, Munn Z, editors. *Joanna Briggs Institute Reviewer's Manual*. : The Joanna Briggs Institute; 2020
105. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing clinical guidelines. *West J Med*. 1999;170(6):348–351.
106. OECD Eurostat and World Health Organization. Classification of Health Care Providers (ICHA-HP). In: *A System of Health Accounts 2011*. OECD Publishing; 2017: 121–152. doi:10.1787/9789264270985-en
107. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021: n71. doi:10.1136/bmj.n71
108. Schreier M. Qualitative content analysis. In: Flick U, editor. *The SAGE Handbook of Qualitative Data Analysis*. Sage Publications; 2014:170–183.
109. World Health Organization. Introduction. In: *WHO Handbook for Guideline Developement*. 2nd ed. World Health Organization; 2014:2
110. Pernambuco L, Espelt A, Magalhães Junior HV, Lima de KC. Recomendações para elaboração, tradução, adaptação transcultural e processo de validação de testes em fonoaudiologia. *CoDAS*. 2017;29(3). doi:10.1590/2317-1782/20172016217
111. International Test Comission. *ITC Guidelines for Translating and Adapting Tests*. International Test Comission; 2017.
112. Tafforeau J, Cobo ML, Tolonen H, Christa S-N, Tinto A. Guidelines for the Development and Criteria for the Adoption of Health Survey Instruments; 2015. Available from: https://ec.europa.eu/health/ph_information/dissemination/reporting/healthsurveys_en.pdf. Accessed May 16, 2024.
113. Høegh MC, Høegh S-M. Trans-adapting outcome measures in rehabilitation: cross-cultural issues. *Neuropsychol Rehabil*. 2009;19(6):955–970. doi:10.1080/09602010902995986
114. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med*. 2011;9(1):86. doi:10.1186/1741-7015-9-86

115. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157. doi:10.1007/s11136-018-1798-3
116. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments; 2019. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed May 16, 2024.
117. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol*. 2010;10(1):13. doi:10.1186/1471-2288-10-13
118. Chávez LM, Canino GC Toolkit on Translating and Adapting Instruments; 2005. Available from: https://www.hsri.org/files/uploads/publications/PN54_Translating_and_Adapting.pdf. Accessed May 16, 2024.
119. World Health Organization. WHO guidelines on translation: process of translation and adaptation of instruments; 2019. Available from: http://www.who.int/substance_abuse/research_tools/translation/en/. Accessed May 16, 2024.
120. Wild D, Eremenco S, Mear I, et al. Multinational trials—Recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the Data: the ISPOR patient-reported outcomes translation and linguistic validation good research practices task force Report. *Value Heal*. 2009;12(4):430–440. doi:10.1111/j.1524-4733.2008.00471.x
121. Swami V, Barron D. Translation and validation of body image instruments: challenges, good practice guidelines, and reporting recommendations for test adaptation. *Body Image*. 2019;31:204–220. doi:10.1016/j.bodyim.2018.08.014
122. Kuliś D, Bottomley A, Velikova G, Greimel E, Koller M. EORTC Quality of Life Group Translation Procedure; 2017. Available from: https://www.eortc.org/app/uploads/sites/2/2018/02/translation_manual_2017.pdf. Accessed May 16, 2024.
123. Ohrbach R, Ohrbach J, Jezewski M, John MT, Lobbezoo F. Guidelines for Establishing Cultural Equivalency of Instruments; 2013. Available from: https://ubwp.buffalo.edu/rdc-tmdinternational/wp-content/uploads/sites/58/2017/01/Guidelines-for-Translation-and-Cultural-Equivalency-of-Instruments-2013_05_118608.pdf. Accessed May 16, 2024.
124. Baker DL, Melnikow J, Ying Ly M, Shoultz J, Niederhauser V, Diaz-Escamilla R. Translation of health surveys using mixed methods. *J Nurs Scholarsh*. 2010;42(4):430–438. doi:10.1111/j.1547-5069.2010.01368.x
125. Biering-Sørensen F, Alexander MS, Burns S, et al. Recommendations for translation and reliability testing of international spinal cord injury data sets. *Spinal Cord*. 2011;49(3):357–360. doi:10.1038/sc.2010.153
126. Koller M, Aaronson NK, Blazeby J, et al. Translation procedures for standardised quality of life questionnaires: the European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur J Cancer*. 2007;43(12):1810–1820. doi:10.1016/j.ejca.2007.05.029
127. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof*. 2005;28(2):212–232. doi:10.1177/0163278705275342
128. Dhamani KA, Richter MS. Translation of research instruments: research processes, pitfalls and challenges. *Afr J Nurs Midwifery*. 2011;13(1):3–13.
129. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology*. 2004;126:S124–S128. doi:10.1053/j.gastro.2003.10.016
130. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193–205. doi:10.1023/a:1015291021312
131. Organisation for Economic Co-operation and Development. Translation and adaptation guidelines for PISA 2012; 2012. Available from: https://www.oecd.org/pisa/pisaproducts/49273486.pdf. Accessed May 16, 2024.
132. Swami V, Todd J, Barron D. Translation and validation of body image instruments: an addendum to Swami and Barron (2019) in the form of frequently asked questions. *Body Image*. 2021;37:214–224. doi:10.1016/j.bodyim.2021.03.002
133. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170. doi:10.1007/s11136-018-1829-0
134. Yasir ASM. Cross Cultural Adaptation & Psychometric Validation of Instruments: step-wise Description. *Int J Psychiatry*. 2016;1(1). doi:10.33140/IJP/01/01/00001
135. Regnault A, Herdman M. Using quantitative methods within the Universalist model framework to explore the cross-cultural equivalence of patient-reported outcome instruments. *Qual Life Res*. 2015;24(1):115–124. doi:10.1007/s11136-014-0722-8
136. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10(1):22. doi:10.1186/1471-2288-10-22
137. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–549. doi:10.1007/s11136-010-9606-8
138. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012
139. Möhler R, Bartoszek G, Köpke S, Meyer G. Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): guideline development. *Int J Nurs Stud*. 2012;49(1):40–46. doi:10.1016/j.ijnurstu.2011.08.003
140. American Educational Research Association. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
141. Gudmundsson E. Guidelines for translating and adapting psychological instruments. *Nord Psychol*. 2009;61(2):29–45. doi:10.1027/1901-2276.61.2.29
142. Forsyth BH, Kudela MS, Levin K, Lawrence D, Willis GB. Methods for translating an English-language survey questionnaire on tobacco use into mandarin, Cantonese, Korean, and Vietnamese. *Field Methods*. 2007;19(3):264–283. doi:10.1177/1525822X07302105
143. American Psychological Association. Criteria for practice guideline development and evaluation. *Am Psychol*. 2002;57(12):1048–1051. doi:10.1037/0003-066X.57.12.1048
144. Iliescu D. Pre-condition guidelines. In: *Adapting Tests in Linguist and Cultural Adaptations*. Cambridge University Press; 2017:87.
145. Iliescu D. Documentation guidelines. In: *Adapting Tests in Linguist and Cultural Adaptations*. Cambridge University Press; 2017:105.
146. Hambletom RK. Issues, designs and tecnical guidelines for adaptating tests into multiples languages and cultures. In: Hembletom RK, Merenda PF, Spilberg CD, editors. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates; 2005:3–38.

147. Oakland T. Selected ethical issues relevant to test adaptations. In: Hambletom RK, Merenda PF, Spielberg CD, editors. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates; 2005:80.
148. Brislin RW, Lonner WJ, Thorndike RM. Cross cultural research methods. In: *Transcultural Psychiatric Research Review*. Vol. 12. John Wiley ans Sons;1973:7–10. doi:10.1177/136346157501200101
149. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Recommendations for the cross-cultural adaptation of the DASH & QuickDASH outcome measure. *Inst Work Heal*. 2007;45.
150. Ozolins U, Hale S, Cheng X, Hyatt A, Schofield P. Translation and back-translation methodology in health research: a critique. *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(1):69–77. doi:10.1080/14737167.2020.1734453
151. Bornman J, Sevcik RA, Romski M, Pae HK. Successfully translating language and culture when adapting assessment measures. *J Policy Pract Intellect Disabil*. 2010;7(2):111–118. doi:10.1111/j.1741-1130.2010.00254.x
152. Simonsen E, Mortensen EL. Difficulties in translation of personality scales. *J Pers Disord*. 1990;4(3):290–296. doi:10.1521/pedi.1990.4.3.290
153. Tanzer NK. *Developing Tests for Use in Multiple Languages and Cultures: A Plea for Simultaneous Development*. Lawrence Erlbaum Associates; 2005.
154. Bennett PM. Reviewing translated scales: backtranslation under the spotlight. *Transl Matters*. 2022;4(125–144).
155. van de Vijver F, Tanzer NK. Bias and equivalence in cross-cultural assessment: an overview. *Eur Rev Appl Psychol*. 2004;54(2):119–135. doi:10.1016/j.erap.2003.12.004
156. de Klerk S, Jerosch-Herold C, Buchanan H, van Niekerk L. Shared decision making and the practice of community translation in presenting a pre-final Afrikaans for the Western Cape Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire: a proposal for improved translation and cross-cultural adaptation. *J Patient-Reported Outcomes*. 2019;3(1):52. doi:10.1186/s41687-019-0144-z
157. Erkut S. Developing multiple language versons of instruments for intercultural research. In: Ji M, editor. *Cross-Cultural Health Translation: Exploring Methodological and Digital Tools*. Routledge; 2019:9–10.
158. Silverblatt A, Zlobin N. Production elements. In: *International Communications: A Media Literacy Approach*. Routledge; 2015:108.
159. St Amant K. A new web for the new millennium. In: Lipson C, Day M, editors. *Technical Communication and the World Wide Web*. Lawrence Erlbaum Associates Publishers; 2005:148.
160. Drasgow F, Probst TM. The psicometrics of adaptation: evaluating measurement equivalance across languages and cultures. In: Hambletom RK, Merenda PF, Spielberg CD, editors. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates Publishers; 2005:361.
161. Erkut S, Alarcón O, Coll CG, Tropp LR, García HAV. The dual-focus approach to creating bilingual measures. *J Cross Cult Psychol*. 1999;30 (2):206–218. doi:10.1177/0022022199030002004
162. van der Vidjer F, Leung K. Equivalence and bias: a review of concepts, models, and data analytic procedures. In: Matsumoto D, van de Vijver F, editors. *Cross-Cultural Research Methods in Psychology*. Cambridge University Press; 2011:17–44.
163. American Educational Research Association. Test design and development. *Standards for Educational and Psychological Testing*; American Educational Research Association; 2014. 81–84.
164. van Teijlingen ER, Hundley V. Pilot study. In: Lewis-Back MS, Bryman A, Liao TF, editors. *The SAGE Encyclopedia of Social Science Research Methods. Vol II*. Sage; 2004:823–824.
165. Gallagher PM. Pretest. In: Lewis-Back MS, Bryman A, Liao TF, editors. *The SAGE Encyclopedia of Social Science Research Methods. Vol II*. Sage Publications; 2004:853–854.
166. Korabik K, van Rhijn T. Best practices in scale translation and establisment mesurement equivalence. In: Shockey KM, Shen W, Johnson RC, editors. *The Cambridge Handbook of the Global Work–Family Interface*. Cambridge University Press; 2018:212–229.
167. Willis G. Pretesting of health survey questionnaires: cognitive interviewing, usability testing, and behavior coding. In: Johnson TP, editor. *Handbook on Health Survey Methods*. Willey; 2015:221.
168. Rodrigues IB, Adachi JD, Beattie KA, MacDermid JC. Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskelet Disord*. 2017;18(1):540. doi:10.1186/s12891-017-1914-5
169. Perneger TV, Courvoisier DS, Hudelson PM, Gayet-Ageron A. Sample size for pre-tests of questionnaires. *Qual Life Res*. 2015;24(1):147–151. doi:10.1007/s11136-014-0752-2
170. Holyk GG. Question testing methods. In: *Encyclopedia of Survey Research Methods. Vol II*. Sage; 2008:658–659.
171. van der Zouwen J, Smit JH, et al. Evaluating survey questions by analyzing patterns of behavior codes and question–answer sequences: a diagnostic approach. In: Presser S, Rothgeb JM, Couper MP, editors. *Methods for Testing and Evaluating Survey Questionnaires*. Wiley Interscience; 2004. 124.
172. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res*. 2003;25(5):508–518. doi:10.1177/0193945903252998
173. Polit DF, Beck CT. Measurement and data quality. In: *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. 9th ed. Wolker Klumer / Lippincott Williams & Wilkins; 2012:334.
174. Linden B, Stuart H. Preliminary analysis of validation evidence for two new scales assessing teachers' confidence and worries related to delivering mental health content in the classroom. *BMC Psychol*. 2019;7(1):32. doi:10.1186/s40359-019-0307-y
175. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: conducting a content validity study in social work research. *Soc Work Res*. 2003;27(2):94–104. doi:10.1093/swr/27.2.94
176. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. *Res Soc Adm Pharm*. 2019;15(2):214–221. doi:10.1016/j.sapharm.2018.03.066
177. DesRoches D. Establishment survey. In: Lavrakas PJ, editor. *Encyclopedia of Survey Research Methods*. Sage; 2008:124.
178. Kelly J, Lavrakas PJ. Debriefing. In: *Encyclopedia of Survey Research Methods. Vol I*. Sage; 2008:181–182.
179. Beatty P. Developing measures of health behavior and he alth service utilization. In: Johnson TP, editor. *Handbook on Health Survey Methods*. Willey; 2015:186–187.
180. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30 (4):459–467. doi:10.1002/nur.20199
181. Yusoff MSB. ABC of content validation and content validity index calculation. *Educ Med J*. 2019;11(2):49–54. doi:10.21315/eimj2019.11.2.6

182. Polit DF, Beck CT. Qualitative research design and approaches. In: *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. 10th ed. Wolters Kluwer; 2017:489.
183. Lawshe CH. A quantitative approach to content validity. *Pers Psychol*. 1975;28(4):563–575. doi:10.1111/j.1744-6570.1975.tb01393.x
184. Ayre C, Scally AJ. Critical values for Lawshe's Content Validity Ratio. *Meas Eval Couns Dev*. 2014;47(1):79–86. doi:10.1177/0748175613513808
185. Stone KS, Frazier SK. Evaluation of measurement precision, accuracy, and error in biophysical data. In: Waltz CF, Strickland OL, Lenz ER, editors. *Measurement in Nursing and Health Research*. 4th ed. Springer Publishing; 2010:387–390.
186. Cicchetti DV. On a model for assessing the security of infantile attachment: issues of observer reliability and validity. *Behav Brain Sci*. 1984;7(1):149–150. doi:10.1017/S0140525X00026558
187. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378–382. doi:10.1037/h0031619
188. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011;48(6):661–671. doi:10.1016/j.ijnurstu.2011.01.016
189. Osborne JW. *Best Practices in Exploratory Factor Analysis*. CreateSpace Independent Publishing; 2014.
190. Lloret-Segura S, Ferreres-Traver A, Hernández-Baeza A, Tomás-Marco I. El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *An Psicol*. 2014;30(3). doi:10.6018/analesps.30.3.199361
191. Field A. Why is my evil lecturer forcing me to learn statistics? In: *Discovering Statistics Using IBM SPSS Statistics*. 5th ed. Sage Publications; 2018:70–71.
192. Kothari CR. Multivariate analysis techniques. In: *Research Methodology: Methods and Techniques*. 2nd ed. New Age Publisher; 2004:322.
193. Haig BD. Exploratory factor analysis, theory generation, and scientific method. In: *Method Matters in Psychology: Essays in Applied Philosophy of Science*. Springer; 2018:65–88.
194. Haig BD. Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behav Res*. 2005;40(3):303–329. doi:10.1207/s15327906mbr4003_2
195. Råholm M-B. Abductive reasoning and the formation of scientific knowledge within nursing research. *Nurs Philos*. 2010;11(4):260–270. doi:10.1111/j.1466-769X.2010.00457.x
196. Prudon P. Confirmatory factor analysis as a tool in research using questionnaires: a critique. *Compr Psychol*. 2015;4:03.CP.4.10. doi:10.2466/03.CP.4.10
197. Eldridge J. Reliability, validity and trustworthiness. In: Boswell C, Cannon S, editors. *Introduction to Nursing Research: Incorporating Evidence-Based Practice*. 5th ed. Jones & Bartlettt Publishers; 2020:271–294.
198. Abu-Bader SH. Working with SPSS. In: *Using Statistical Methods in Social Science Research: With a Complete SPS Guide*. 2nd ed. Oxford University Press; 2021:56.
199. Perrin KM. Reliability and validity. In: *Principles of Planning, Evaluation and Research for Health Care Programmes*. 2nd ed. Jones &: Bartlett Learning; 2022:141.
200. Grove SK. Quantitative measurement concepts. In: Gray JR, Grove SK, editors. *Burns & Groves the Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence*. 9th ed. Elsevier; 2021:462.
201. Hair J, Hollingsworth CL, Randolph AB, Chong AYL. An updated and expanded assessment of PLS-SEM in information systems research. *Ind Manag Data Syst*. 2017;117(3):442–458. doi:10.1108/IMDS-04-2016-0130
202. Field A. Exploratory factor analysis. In: *Discovering Statistics Using IBM SPSS Statistics (North American Edition)*. 5th ed. Sage; 2018:1390–1391.
203. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. *Psychol Assess*. 2019;31(12):1412–1427. doi:10.1037/pas0000626
204. Moreira J. *Questionários: Teoria e Prática*. Almedina; 2004.
205. Streiner DL, Norman GR, Cainey J. Basic concepts. In: *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. Oxford University Press; 2015:8.
206. Bryant F. Assessing the validity of measurement. In: Grimm LG, Yarnold PR, editors. *Reading and Understanding More Multivariate Statistics*. American Psychological Association; 2002:99–146.
207. Marôco J, Etapas da análise de equações estruturais. 3rd. *Análise de Equações Estruturais*, 2021:55
208. Sommer I, Larsen K, Nielsen CM, Stenholt BV, Bjørk IT. Improving clinical nurses' development of supervision skills through an action learning approach. *Nurs Res Pract*. 2020;2020:1–10. doi:10.1155/2020/9483549
209. Bowling A. Quality of life: concepts, measurements and patient perception. In: *Research Methods in Health: Investigating Health and Health Services*. 4th ed. Open University Press; 2014:54.
210. Bot SM, Terwee CB, van der Windt DAWM, Boute LM, Dekker J, Vet de HCW. Psychometric evaluation of self-report questionnaires: the development of a checklist. In: Adér HJ, Mellenberg C, editors. *Proceedings of the Second Workshop on Research Methodology.* Vu University; 2003:161–168.
211. Powers BA, Knapp TR. Face validity. In: *Dictionary of Nursing Theory and Research*. 3rd ed. Spinger Publishing; 2006:63.
212. Soeken WKL. Validity of measures. In: *Measurement in Nursing and Health Research*. 4th ed. Springer Publishing Company; 2010:163.
213. Streiner DL, Norman GR, Cairney J. Seleting the itens. In: *Health Measurement Scales: A Practical Guide to Their Development and Use. 5th*. Oxford University Press; 2015:92.
214. Zumbo BD. Validity as contextualized and pragmatic explanation, and its implications for validation practice. In: Lissitz RW, editor. *The Concept of Validity: Revisions, New Directions, and Applications*. Information Age Publishing; 2009:68.
215. 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*. Springer Publishing; 2014:9–24.
216. Mokkink LB, Terwee CB, Knol DL, et al. Taxonomy and definitions. *COSMIN Checklist Manual*. 2012;9.