




# The development of the Quality of Translation and Linguistic Equivalence Checklist



## Authors:

Mario R. Smith<sup>1</sup> 

Nuraan Adams<sup>1</sup> 

Erica Munnik<sup>1</sup> 

## Affiliations:

<sup>1</sup>Department of Psychology, Faculty of Community and Health Sciences, University of the Western Cape, Cape Town, South Africa

## Corresponding author:

Erica Munnik,  
emunnik@uwc.ac.za

## Dates:

Received: 22 Feb. 2022

Accepted: 13 July 2022

Published: 26 Oct. 2022

## How to cite this article:

Smith, M.R., Adams, N., & Munnik, E. (2022). The development of the Quality of Translation and Linguistic Equivalence Checklist. *African Journal of Psychological Assessment*, 4(0), a108. <https://doi.org/10.4102/ajopa.v4i0.108>

## Copyright:

© 2022. The Authors.  
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

The International Test Commission (ITC) established guidelines for test adaptations. The ITC encourages the adaptation of locally developed measures with proven validity. A good quality translation process ensures that the same meaning is conveyed from the source to the target language. Through test adaptation, researchers focus on cultural differences between the source and the target language to maintain linguistic equivalence. Research involving adaptation has systematically failed to report on the rigour of the translation process and to make translation part of the empirical process. The ITC guidelines are generally referred to; however, the assessment of the quality of translations and the process of establishing linguistic equivalence remain an important research focus. This study reports on the development of the Quality of Translation and Linguistic Equivalence Checklist (QTLC). The construction of the QTLC was based on ITC guidelines. The QTLC consists of two sections, translation and linguistic equivalence, and produced section scores with accompanying quality descriptions. The draft instrument was presented to three independent reviewers. Once feedback was incorporated, the QTLC was piloted in an ongoing study on the translation of the E3SR. Two reviewers applied the checklist, and inter-rater reliability was established. The Kappa statistic (0.78) tested significant at a 0.00 alpha level, indicating substantial agreement between the raters on the quality of the translation process and equivalence. Four items were identified as functioning differently and were subsequently revised. The QTLC appears to be a robust checklist assessing the quality of translations and the process of establishing linguistic equivalence.

**Keywords:** adaptation; linguistic equivalence; inter-rater reliability; translation; QTLC.

## Introduction

Access to reliable and valid measures is critical to the provision of culture-fair quality assessment practices (Sousa & Rojjanasrirat, 2011). In a multilingual country such as South Africa, culture-fair assessment practices require reliable and valid assessment measures that are available in different languages (Mohamed, 2013). South Africa has 11 official languages, but services and instrumentation are largely available in English and Afrikaans (Munnik, Wagener, & Smith, 2021). Hernández, Hidalgo, Hambleton and Gomez-Benito (2020) recommended that locally developed measures with proven validity should be considered for translation, as they already have a higher level of contextual relevance than measures developed in other countries or cultures. Thus, the need for the translation of measures, especially locally developed measures, is an important area of research currently.

Hernández et al. (2020) contended that translation is only one part of a multifaceted process called test adaptation. Test adaptation attempts to modify the content of an instrument to make it culturally appropriate and accurate (Epstein, Santo, & Guillemin, 2015). This process includes translation, equivalence and validation (Lakens, Scheel, & Isager, 2018). Translation entails the written rendering of the meaning of a word from the source language to the target language. The translated version must be as close as possible to the format of the original and must consider possible linguistic challenges. The quality of the translation process ensures that the same meaning is conveyed from the source language to the target language (Behr, 2017).

Linguistic equivalence ensures that there is similarity in meaning between two sets of words spoken or written in different languages (Geisinger, 2003). In other words, the translation of an item from the source language to the target language will convey the same meaning. Researchers must adjust the content to avoid culturally biased items and poor phrasing whilst ensuring that the content in the translated version is comparable to the original. In this way, they are attempting to maintain linguistic equivalence (Arnold & Smith, 2013).

## Read online:



Scan this QR code with your smart phone or mobile device to read online.

Rawoot and Florence (2017) stated that the process of adaptation usually requires well-defined and executed steps. Similarly, Lakens (2017) recommended that adaptation requires careful planning and must follow a rigorous and comprehensive empirical process. To this end, Hambleton (2011) recommended the guidelines proposed by the International Test Commission (ITC) (ITC, 2016). The ITC comprises various psychological associations, publishers, test commissions and other organisations committed to promote effective testing and assessment policies in the construction and evaluation of instruments through the guidelines that they develop (ITC, 2016). A detailed account of the guidelines is provided under the heading conceptual framework.

Hambleton (2011) criticised test adaptation practices and underscored that researchers do not consistently and systematically report on the methodological rigour and coherence of the translation process and whether or how linguistic equivalence is achieved. Researchers focus on data reduction techniques to demonstrate construct validity and Cronbach's alpha to demonstrate internal consistency of the translated measures (Peters, 2014; Tavakol & Dennick, 2011). Cross-cultural test constructors remain concerned about the lack of systematic methodology and quality assurance of test adaptation through translation (Arafat, Chowdhury, Qusar, & Hafez, 2016). The guidelines proposed by the ITC are often cited or referred to without an account of how it was applied (Hernández et al., 2020). Similarly, there is no existing means to evaluate and report on translation processes against the ITC guidelines. To address this identified gap in the literature, the authors developed and piloted a checklist based on the ITC guidelines against which translation processes can be assessed. This manuscript reports on the development of the quality of translation and linguistic equivalence checklist (QTLC).

## Conceptual framework

The second edition of the ITC Test Translation and Adaptation Guidelines was adopted as the conceptual framework for the study (ITC, 2016). The guidelines were developed as the adaptation processes in research did not always follow a rigorous process (ITC, 2016). The ITC guidelines consists of general guidelines that constitute a framework for good practice in test adaptation. The ITC guidelines are structured into four divisions for ease of use, namely (1) precondition, (2) test development and confirmation, (3) administration and (4) documentation. Each will be discussed in turn.

1. 'Precondition' emphasises that decisions must be made before the adaptation process begins. An important consideration at this stage is whether the construct has an equivalent in the target language, that is, linguistic equivalence. The guidelines recommend that permission needs to be obtained from the test developer. The amount of overlap in the definition and content of the construct being measured must be estimated. In addition, the item content in the populations of interest must be

assessed to determine if it is sufficient for the intended use(s) of the instrument.

2. 'Test development and adaptation' focuses on the actual process of adaptation. Adaptation is a more expansive term referring to the conversion of content in one language and culture into another language. Adaptation refers to activities that include deciding whether or not a test or instrument in a second language and culture could measure the same construct as it was intended to measure in the first language. Adaptation includes translation that entails converting the content of a test from one language to another to preserve the linguistic meaning and establishing whether the resultant translation is equivalent to the version in the source language (ITC, 2016). The guidelines recommend the use of appropriate translation designs and procedures to maximise the suitability of the test adaptation in the intended populations.

Translators must be selected carefully and should have demonstrable expertise in translation over and above fluency in the source and target languages (ITC, 2016). Expert knowledge of the subject matter is also recommended (Graneheim & Lundman, 2004). When conducting the translation, translators must complete their work without prior knowledge of the instrument and of each other (Odero, 2017). Two independent sets of translators must complete the forward and backward translations. A minimum of two translators are recommended per translation (ITC, 2016).

A design for evaluating the work of test translators must be selected. The guidelines recommend the use of forward and backward translations as an acceptable method for evaluating the quality of translations (Chidlow, Plakoyiannaki, & Welch, 2014). It is also recommended that a team implements a formal evaluation process that is ideally audited externally (Odero, 2017). The evaluation process must consider any necessary accommodations including modification of the test format and revisions to the source format if it enhances the meaning (Odero, 2017). An important focus in this section is to ensure that the translation and adaptation processes consider linguistic and cultural differences in the intended populations.

This section further focuses on establishing linguistic equivalence between the test in the source and target languages and cultures. The guidelines recommend that evidence is provided to confirm that test instructions and item content have similar meaning for all intended populations. The item formats, rating scales, scoring categories, test conventions, modes of administration and other procedures must also be suitable for all intended populations (ITC, 2016).

Various forms of evidence have been suggested. Geisinger (2003) asserted that linguistic equivalence can be achieved in two possible ways. Firstly, equivalence can be established through high-quality back translation. Back translation entails translating the target language back to the source

language independently to ensure that the target language carries that same meaning as the source language (Chen & Boore, 2010). Cash and Snider (2014) recommended that both translators should be bilingual speakers and knowledgeable of the topic under study to ensure that equivalence is maintained. Secondly, manifest and latent content analysis can be used to establish linguistic equivalence (Chen & Boore, 2010). Graneheim and Lundman (2004) described manifest content analysis that focuses on the content aspect and components within a text, whereas latent content analysis is involved with the underlying meanings of interpretations. Manifest content analysis is more objective in nature, whereas latent content analysis is more subjective. Omar (2012) highlighted the importance of understanding the use of these concepts in context as it influences the grammatical, semantic, social and cultural meanings. Small pilot studies are recommended. Pilot studies provide data generated by the adapted instrument that can be subjected to techniques such as item analysis, reliability assessment and small-scale validity studies (Hernández et al., 2020). The results of these studies can inform any necessary revisions to the adapted test.

The guidelines recommend that the initial evidence is followed up with full-scale (larger) validity studies (ITC, 2016). The analysis in such studies provide relevant statistical evidence about construct equivalence, method equivalence and item equivalence for all intended populations. This process also provides evidence supporting the norms, reliability and validity of the adapted version of the test in the intended populations. This part of the guidelines is referred to as 'confirmation' and includes the gathering of empirical evidence to address the equivalence, reliability and validity of a test or instrument in multiple languages and cultures (Hambleton, 2011).

- Administration relates to 'administration' and 'score scales and interpretation'. This set of guidelines aims to provide direction as to how the assessment should be administered in the different languages. It underscores the importance of providing clear instructions so that the test or instrument is used as it is intended (ITC, 2016). Variation in instrumentation (how the test is used) can invalidate the resulting profile. Thus, clear and explicit written instructions should be provided for each version.
- Documentation has been a particularly neglected topic in test adaptation (Hambleton, 2011). The guidelines demand more when it comes to documentation of the test adaptation process (ITC, 2016). The focus in this section is to prepare supplementary materials and instructions to minimise any culture- and language-related problems that are caused by administration procedures and response modes that can affect the validity of the inferences drawn from the scores. It also includes the specification of any testing conditions that should be followed closely in all populations of interest. The provision of technical documentation to note any changes, including an account

of the evidence obtained to support equivalence, when a test is adapted for use in another population is emphasised (Hambleton, 2011). This is important to enable good practice amongst test users using the adapted test with people in the context of the new population. Documentation also extends to the publication of information in the form of manuals. Dissemination through journal publications should specifically report on the methodological rigour and coherence of the adaptation process (Hambleton, 2011). Similarly, funding instruments should request information about the planned documentation or dissemination protocol in order to enhance knowledge translation for test users and science communication to test constructors and test adaptors (Hambleton, 2011).

The conceptual framework informed the overall aim of the study, which was to develop a checklist for assessing the quality of translation and equivalence processes. The ITC guidelines formed the theoretical underpinning of the proposed checklist and also informed subsequent methodological decisions.

## Aim of the study

The aim of the study was to design and develop the QTLC that can evaluate the quality of processes used in test translation and in the establishment of linguistic equivalence.

## Methods

### Design

This construction study consisted of two phases. Phase 1 entailed the construction of the QTLC. Phase 2 entailed piloting of the checklist.

#### Phase 1

The construction followed a five-step process. The first step entailed selecting a theoretical structure for the checklist based on the ITC guidelines for test adaptation (ITC, 2016). The second step entailed deciding on the format of the checklist and the quantification for scoring purposes. The third step entailed generating a pool of items and finalising the draft checklist. The fourth step entailed reviewing and refining the draft scale. The fifth step entailed developing the accompanying templates and instruction guide. The steps are elaborated as follows.

**Step 1: Theoretical structure:** As mentioned before, the conceptual framework formed the theoretical basis of the proposed measure. The ITC guidelines for test adaptation formed the primary theoretical tenets that underpin the proposed measure. (ITC, 2016). Thus, the ITC guidelines for adaptation through translation were defined for measurement, that is, operationalised. The resultant measure is called the QTLC (Appendix 1).

**Step 2: Format of the instrument:** The checklist format was deemed appropriate as it would allow using the ITC

guidelines as the basis for items. Each item corresponds to criteria recommended by the ITC for good practice in translation and establishing linguistic equivalence. The checklist was divided into two sections to address the processes for translation and linguistic equivalence respectively.

Section 1 deals with translation and contains two subsections. Subsection 1 deals with the experience of translators, their formal qualification and cumulative experience of the translators. Subsection 2 relates to the process of translation.

Section 2 deals with linguistic equivalence and has three subsections. Subsection 1 addresses the comparison between the original (source document) and draft in the target language. Subsection 2 assesses the comparison between the translated version and back translations. Subsection 3 evaluates the comparison between the original version (source document) and back-translated drafts. The three subsections reflect the assumption that good practice would include forward and back translation, as well as comparisons between the different versions produced.

A sliding scale was adopted for quantification and scoring purposes. Each item is scored, where higher scores indicate a higher-quality response. It was decided that each subsection would generate a score that is the sum of the scores on items in that subsection. Each section produces a section score, which is the sum of the subsection scores. This structure was based on the recommendations of Mahmood and Jacobo (2019), in which the scoring is best understood as a cumulative process, producing scores that can be interpreted independently for subsections and cumulatively for composite scores.

An interpretation matrix was designed to assist with interpretation of the scores (Appendix 2). Each section was assigned a quality description that guides the interpretation of the composite (section) scores. Quality descriptions for Section 1 described the quality of the translation. Descriptions for Section 2 describe the quality of the process for establishing equivalence. Three quality descriptions were distilled, namely (1) poor, (2) good and (3) excellent. Each quality description or category had corresponding actions that can assist researchers or test constructors to apply corrective actions. Scores were expressed as a percentage to guide the quality descriptions. Poor compliance is considered to be reflective of scores below the 50% threshold. Good compliance was considered to be reflective of scores ranging between 50% and 79%. Excellent compliance was considered to be reflective of scores equal to or exceeding 80%.

As mentioned here, Section 2 consisted of three subsections that evaluate independent aspects or processes. Thus, it was decided to apply the quality descriptors to the subsections in Section 2 as well. This enables the instrument to be used in a formative manner when assessing the process followed to establish linguistic equivalence.

**Step 3: Item generation:** For Section 1, items were formulated that assessed the formal qualification and cumulative experience of the translators, the number of translators involved, the process of comparing different versions of the translations, whether back translation was conducted and how an integrated final version was produced. Items were generated for each of the criteria stipulated in the ITC guidelines. Items were not generated for the precondition. Obtaining permission to use an instrument for adaptation was considered an ethics principle. Thus, this particular precondition could be assessed under ethics. The guidelines pertaining to the existence of an equivalent in the target language was considered an important aspect in the rationale for pursuing an adaptation study including translation and linguistic equivalence. As such, this can be assessed under the rationale for an adaptation study. The inclusion of items from guidelines about preconditions was considered necessary but not sufficient to evaluate the quality of translation and equivalence processes. Thus, these can be addressed relatively easily as mentioned here, and their inclusion in a checklist was thought not to add much value to the assessment of quality.

For Section 2, the items were generated to assess whether the meaning of items was captured accurately. Items across all three subsections aimed to evaluate the manifest and latent content of translated items in terms of clarity and lack of ambiguity. Items across this section assess whether the meaning of items were accurately captured. The draft checklist consisted of 37 items. Section 1 included 16 items. Section 2 included 21 items with seven items per subsection.

**Step 4: Reviewing and refining the draft scale:** The draft checklist was reviewed by two independent reviewers who were registered research psychologists ( $n = 2$ ) with the Health Professions Council of South Africa (HPCSA). The reviewers had expertise in research methodology, psychometric test construction and psychological assessment, as evidenced by their qualifications, work experience and publications in the areas mentioned. The reviewers identified that the flow of items in Section 2 was confusing, and items seemed to be repetitive. The items ( $n = 7$ ) were revised to create a better progression and to remove the appearance of repetition. The reviewers also indicated that the composite scores for sections and subsections were higher than the maximum score indicated. This was revised accordingly, but it did not impact the number of items or the structure of the draft checklist.

Each item is scored using a sliding scale, where higher scores indicate a higher quality response. Each section and subsection generates a score that is summed across the items in that section or subsection. The scoring grid was finalised. Section 1 produces a maximum composite section score of 32 that comprises the Subsection 1 score (a maximum score of 18) and Subsection 2 score (a maximum score of 14). Section 2 produces a maximum composite section score of 39 that comprises the Subsection 1 score (a maximum score of 13), Subsection 2 score (a maximum score of 13) and Subsection 3 score (a maximum score of 13).

*Interpretation:* Each section has a quality description that guides the interpretation of the subsections and composite scores. Three quality assurance descriptions were defined, namely (1) poor, (2) good and (3) excellent. Quality descriptions for Section 1 describe the quality of the translation. Descriptions for Section 2 describe the quality of the process for establishing linguistic equivalence. Each quality description or category has corresponding corrective actions that can be undertaken by the researcher.

**Step 5: Developing accompanying templates and instruction guide:** Two accompanying documents were compiled and included in the QTLC. The first document is the QTLC template that the researcher(s) responsible for the translation completes (Appendix 3). This template corresponds to the items and sections of the checklist. The researcher(s) captures the details of the translation and equivalence processes on this template. The template is used by reviewers as the source document for their evaluations. The motivation for this template was to create a higher level of consistency and uniformity in presenting the content on the adaptation process. It also reduces bias where researchers familiar with the checklist may be advantaged and can tailor the presentation of their information.

The second document is the reviewer response form (Appendix 4). The response form includes the items and scoring options. Reviewers are the intended users of the response form. This response form facilitates ease of use, as provision is made for the reviewer to record his or her scores.

An interpretation matrix was designed that included a guide to interpretation in tabular form. The tables contain the categorisation of composite scores and the corresponding quality description and corrective actions. For translation, scores below 50% (less than 16 out of a possible 32) indicate a low level of compliance with ITC guidelines and are given a 'poor' rating. Researchers are recommended to redo the translation as per the recommended guidelines in such cases. Scores between 50% and 79% (between 16 and 24 out of a possible 32) were given a rating of 'good'. Such a quality description indicates that there was basic compliance with the guidelines. Researchers are recommended to identify and revise items where concerns were raised. Scores above 79% (25 or more out of a possible 32) were given a quality description as 'Excellent'. This indicates that there was a high level of compliance with the ITC guidelines.

For linguistic equivalence, scores below 50% (less than 19 out of a possible 39) indicate a low level of compliance with the guidelines. Researchers are recommended to redo the equivalence process in compliance with the recommended guidelines. Scores between 50% and 79% (between 19 and 30 out of a possible 39) indicate a basic level of compliance. Researchers must identify and revise items or subsections where concerns have been raised. Scores equal to and above 80% (between 31 and 39) suggest a high level of compliance and were given a quality description as 'excellent'. Based on

this evaluation, researchers can reasonably conclude that linguistic equivalence was achieved. Given the nature of linguistic equivalence, the quality description was also applied to the subsections to assist researchers to identify areas where they can improve or enhance the process.

## Phase 2

Piloting entailed an application of the instrument to the translation process of the Emotional-Social Screening tool for School Readiness (E3SR) from English into Afrikaans. The E3SR is a locally developed screening instrument that assesses preschoolers' emotional and social competencies before entry into mainstream education. It has six factors: emotional maturity, emotional management, sense of self, readiness to learn, social skills and communication. Research on the E3SR established construct validity and reliability (Munnik et al., 2021). More recently, research on the E3SR focused on its translation into Afrikaans.

Additional considerations were that the construct 'emotional-social competence' had an equivalent in the target language, Afrikaans (Bornman & Potgieter, 2017). Afrikaans has been well established as an academic language and has been used widely in education in South Africa. Therefore, equivalent constructs were readily available for most psychological constructs, including emotional-social competence. There was a clear indication from experts in development and education that the definition and content of the construct 'emotional-social competence' was well defined in the source language with established equivalents in the target language (Munnik & Smith, 2019). Therefore, the Afrikaans lexicon or vocabulary sufficiently covered the denotations and connotations of the content (constructs, domains and attributes) that required translation. As a result, translated items could be developed that were appropriate for use with the intended population (preschoolers) and Afrikaans-speaking respondents who would complete the screening tool. These considerations were aligned well with the preconditions outlined in the ITC guidelines. The translation and adaptation of the E3SR from English to Afrikaans was deemed appropriate for piloting the QTLC.

## Translation of the emotional-social screening tool for school readiness

The translation of the E3SR followed the operational steps proposed by Sousa and Rojjanasrirat (2011) as follows.

### Step 1: Translation of the original emotional-social screening tool for school readiness into Afrikaans

The E3SR was translated from English to Afrikaans by two independent translators. The translators were fluent in English and Afrikaans. Translator 1 was a clinical psychologist with 45 years of focused experience in translation and editing. Translator 2 was a research psychologist with expertise in test construction and psychometrics and possessed 40 years of experience in translation. This step generated two translations, independently labelled as TL-1 and TL-2.

### Step 2: Comparison of the two translated versions (TL-1 and TL-2)

The two forward-translated versions of the instrument (TL-1 and TL-2) were compared for ambiguities and discrepancies by the two authors, both qualified psychologists with experience in test development and the content domain. Differences were discussed and resolved by the research team. This step resulted in a final draft (TL-3). An external auditing process was conducted by the third author to distil a final translation.

### Step 3: Back translation of the initial translated version

The translated version (TL-3) was translated back into English by three independent translators who produced three back-translated versions (B-TL1, B-TL2 and B-TL3). The second set of translators had no prior knowledge of the original draft and performed their translations blind. Translator 1 was a clinical psychologist with expertise in clinical practice, language studies and 4 years of translation experience. Translator 2 was a research psychologist with expertise in research methodology, building capacity and qualifications in both editing and language studies, with 3 years of translation experience. Translator 3 was a linguist with expertise in language, education, communication studies and translation and had 30 years of translation experience.

### Step 4: Comparison of the back translated versions (B-TL1, B-TL2 and B-TL3)

The back translations were compared with the original E3SR for format, wording, grammatical structure and meaning by the two authors. Ambiguities and discrepancies regarding cultural meaning and colloquialisms, idioms in words, sentences between back translations and the original E3SR were discussed and resolved between the researchers and the translators. An external auditing process was conducted by the third author to distil a final translation.

### Step 5: Assessing the quality of the translation process and establishing linguistic equivalence

The QTLC was piloted during this step. Two independent reviewers assessed the quality of the process using the QTLC. Reviewer 1 (R1) was a research psychologist who had expertise in the field of statistical techniques and psychometric test construction. Reviewer 2 (R2) was a research psychologist with expertise in capacity development and transferable skills training in research methodology. The reviewers submitted their reviews of the E3SR, and their scores were entered into a composite sheet for ease of comparison and the calculation of inter-rater reliability. The reviewers were also asked to provide qualitative feedback on the QTLC. The comments of the reviewers were tabularised and presented as Appendix 5. The table includes general comments on the QTLC and comments on specific items.

## Procedure and data analysis

The details of the translation and equivalence processes in the translation of the E3SR were recorded on the QTLC

template. This populated template was given to the reviewers as the source document for the evaluation. The reviewers used the reviewer response form to record their scoring of the items and tallying of subsection and section scores.

### Inter-rater reliability

The Kappa statistic was used to calculate inter-rater reliability. The Kappa statistic uses cross-tabulations to assess inter-rater reliability (Field, 2013). A threshold Kappa statistic of 0.61 was established, which is described as a substantial agreement by Glen (2014). The inter-rater reliability provided evidence on the agreement between raters when using the QTLC to assess the quality of the processes followed in the translation of the E3SR.

### Ethical considerations

Ethical clearance was obtained from the Human and Social Science Research Ethics Committee (reference number: HS21/9/2) at the University of Western Cape. Permission was given by Dr Munnik to use the translation of the E3SR for piloting of the QTLC. All personal data of translators and reviewers were de-identified and stored in line with the specified guidelines of the *Protection of Personal Information Act* No. 1 of 2019 (POPIA). Translators and raters signed a binding agreement to maintain independence of their contributions. The agreement included an undertaking to uphold any copyright and intellectual property stipulations by the authors of the E3SR, and the QTLC.

## Results

### Phase 1: Construction

The draft QTLC was found to be clear and coherent by the reviewers. Specific comments on the scoring of individual items were raised. These recommendations were applied that resulted in a simplified and more unified scoring grid. In particular, items asking about the experience of translators were revised to list translators separately and evaluate them separately. A weighted score was introduced for these items, which are described in the scoring section. The reviewers reported that the addition of the QTLC template was crucial and that the researchers had to take responsibility to complete this in a detailed manner, as it formed the source document for reviewing the adaptation processes. The alignment of the structure of the QTLC template and the reviewer report form to that of the QTLC was found to be very helpful.

### Phase 2: Piloting

The results are presented per section of the QTLC for ease of presentation.

#### Section A: Rating the translation processes

The raters scored Subsection 1 identically. The raters awarded a subsection score of eight out of nine, indicating that the translators involved had a high level of experience relevant to translation in the source and target languages. Both raters

awarded a score of 16 for Subsection 2, which was the maximum score possible. The section score was the sum of the two subsection scores. A section score of 24 was attained on both ratings. The corresponding quality description indicated that a high level of compliance with the ITC guidelines was achieved in this translation process. The recommended action was to proceed to establish linguistic equivalence.

### Section B: Rating linguistic equivalence

The scoring in Subsection 1 was identical for both raters. A subsection score of 15 was attained. This score rates the equivalence achieved between the original and the resultant Afrikaans version (TL-3) as excellent, as evidenced by a score exceeding 11.

The scores of the raters for Subsection 2 differed by two points. Rater 1 assigned 14 points, whereas Rater 2 awarded 12 points. The difference was on the items that dealt with the resolution of differences. The text on the QTLC template indicated that the reviewers discussed the differences and reached a decision. Rater 1 interpreted this to mean 'consensus' and scored three points. Rater 2 interpreted this as resolution by 'discussion' and scored two points. The scores indicated that excellent adherence to the ITC guidelines for equivalence was achieved between the Afrikaans E3SR and back translations, as evidenced by a score exceeding 11.

There was a difference of two points between the ratings awarded for Subsection 3. As before, the difference occurred on two items dealing with how differences were resolved. The QTLC does not make the distinction between discussion and consensus clear, resulting in the different interpretation of the reviewers on these items. The scores, 14 and 12, respectively, indicated that excellent equivalence was achieved between the English E3SR and back translations.

The section score was the sum of Subsections 1, 2 and 3. The section scores awarded by Rater 1 (43) were four points higher than Rater 2 (39). Both scores indicate that a high level of compliance to the ITC guidelines for establishing equivalence between the English and the Afrikaans versions of the E3SR was achieved. Linguistic equivalence between the English and Afrikaans draft has therefore been endorsed.

The Kappa statistic (0.78) tested significant at a 0.00 alpha level. There was a substantial agreement between the raters on the quality of the translation process and equivalence. High inter-rater reliability was achieved, despite the response options being interpreted differently by the reviewers. Revisions of the scoring on the identified items and response options was addressed in a subsequent revision.

## Discussion

The ITC guidelines for adaptation through translation and linguistic equivalence are established and widely accepted. However, the lack of a formal checklist hampered the

systematic application in adaptation studies. Similarly, systematic reporting was lacking. The construction of the QTLC addressed an important gap in the body of literature. The checklist format was easy to administer. The processes followed in the construction of the QTLC followed a systematic process and demonstrated a high level of alignment with the ITC guidelines that deal specifically with test adaptation. As mentioned here, the QTLC excluded guidelines related to the preconditions, as these are thought to be covered in general research processes and reporting conventions. The checklist is formative because it identifies areas where there may be concerns about the level of compliance with the ITC guidelines. The interpretation of scores includes a useful recommendation for corrective action that can enhance the processes. The resultant checklist constituted an operationalisation of the ITC guidelines for good practice in translation and establishing linguistic equivalence.

The response options on two items in Subsection 2 and two items in Subsection 3 were interpreted differently by the reviewers. The lack of a clear distinction between the terms 'discussion' and 'consensus' as means of resolving differences resulted in raters scoring differently. This limitation was offset by follow-up discussions with the raters to understand the reasoning behind the difference in their scoring. The respective section scores still attained the same quality description. Thus, the difference in scoring impacted the section score quantitatively but not the corresponding quality description. The revision of the affected items is a priority in further refinement.

The QTLC was successfully used to evaluate the translation of the E3SR into Afrikaans. The finding suggests a high level of compliance with the ITC guidelines in the processes followed during translation and linguistic equivalence between the resultant Afrikaans translation and the original English version of the E3SR. The excellent rating obtained provides a basis for concluding that the resultant translation was linguistically equivalent to the original English version of the E3SR.

The following limitations were observed: the QTLC was only piloted in one translation study. The findings, although encouraging, need to be replicated in more studies. The theoretical underpinnings of the instrument are closely aligned with the ITC guidelines. Thus, the interpretation of the QTLC must be performed in relation to the ITC guidelines and it may not reflect other criteria contained in guidelines that were developed separately.

The item assessing the experience of the translators was scored based on the cumulative experience of the translators. During the review period, this item was flagged as problematic, as it might not accurately reflect differences in experiences between examiners. In the template, the researcher is required to record the exact experience in years. The scoring grid was retained as per recommendations of reviewers. Scoring for this item was amended to score

translators separately. In addition, provision was made for assessing the translators for forward and backward translation separately resulting in two items.

Provision was made for additional translators to be included. Increasing the number of translators above two could exceed the threshold criterion, and therefore the cumulative maximum score could increase correspondingly. This challenge was addressed by introducing a weighted score. The weighted score also made it possible for all translators to be included and evaluated separately. This avoided aggregated scores masking the differences between translators.

The maximum score for each of these items was based on a threshold of two translators who both would have the highest level of experience. The maximum score for forward translation and back translation would be 6 (2 translators  $\times$  3 points), respectively. The maximum scores for these items that can be added to the subsection and section scores was set at 6. This was based on the threshold expectation of two translators with the maximum score awarded for experience (2  $\times$  3 = 6). In other words, above threshold practices would not result in an inflation of the maximum score and the overall section score. The subsection score increased from 9 to 18 and the section score from 23 to 32. Quality descriptions are still based on the stated percentages, but the value of scale scores would increase for section one. The formula to calculate the weighted score is included on the QTLC and the reviewer form.

## Implications for future research, practice and theory

The QTLC attempts to distil the guidelines proposed by the ITC for the processes for translation of instruments and establishing linguistic equivalence. This checklist creates a means for empirically evaluating the translation process from a theory-driven perspective that produces quantifiable outcomes. The checklist contributes to making the methodology underpinning translation explicit, which improves upon the tacit and implicit assumptions offered in the reporting of adaptation studies. The adoption of the QTLC through reuse provides an avenue for making the translation process part of the methodology of adaptation studies and centralise translation and equivalence as a core aspect of the adaptation process.

## Conclusion

The QTLC is a robust checklist that is conceptually grounded in the globally accepted ITC guidelines. This checklist provides a quantifiable methodology for assessing the quality of the processes followed in translation and the establishment of linguistic equivalence. The QTLC provides a method for making implicit processes explicit that in turn enhances the quality of reporting on adaptation through translation and equivalence.

## Acknowledgements

The reviewers involved in the piloting of the checklist are hereby acknowledged for their constructive feedback and contribution to the study.

## Competing interests

The authors confirm that there are no financial or personal relationships that may have improperly influenced them in writing this article.

## Authors' contributions

M.R.S. developed the QTLC and conceptualised the article. M.R.S. contributed to the writing of the article. N.A. piloted the instrument as part of her research towards a postgraduate qualification. The author contributed to the writing of the article. E.M. contributed to the review of the checklist and contributed to the piloting and revision of the checklist. This author contributed to the writing of the article and acted as the corresponding author.

## Funding information

The National Research Foundation (NRF) provided financial assistance through the Thuthuka instrument to the first author. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

## Data availability

The data that support the findings of this study can be made available by the corresponding author, E.M., upon reasonable request.

## Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

## References

- Arafat, S.Y., Chowdhury, H.R., Qusar, M.M.A.S., & Hafez, M.A. (2016). Cross cultural adaptation and psychometric validation of research instruments: A methodological review. *Journal of Behavioural Health*, 5(3), 129–136. <https://doi.org/10.5455/jbh.20160615121755>
- Arnold, B.R., & Smith, J.L. (2013). Methodologies for test translation and cultural equivalence. In F. Paniagua & A. Yamada (Eds.), *Handbook of Multicultural Mental Health* (pp. 243–262). San Diego: Elsevier, Academic press.
- Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, 20(6), 573–584. <https://doi.org/10.1080/13645579.2016.1252188>
- Bornman, E., & Potgieter, P.H. (2017). Language choices and identity in higher education: Afrikaans-speaking students at UNISA. *Studies in Higher Education*, 42(8), 1474–1487. <https://doi.org/10.1080/03075079.2015.1104660>
- Cash, P., & Snider, C. (2014). Investigating design: A comparison of manifest and latent approaches. *Design Studies*, 35(5), 441–472. <https://doi.org/10.1016/j.destud.2014.02.005>
- Chen, H.Y., & Boore, J.R. (2010). Translation and back-translation in qualitative nursing research: Methodological review. *Journal of Clinical Nursing*, 19(1–2), 234–239. <https://doi.org/10.1111/j.1365-2702.2009.02896.x>
- Chidlow, A., Plakoyiannaki, E., & Welch, C. (2014). Translation in cross-language international business research: Beyond equivalence. *Journal of International Business Studies*, 45(5), 562–582. <https://doi.org/10.1057/jibs.2013.67>



- Epstein, J., Santo, R.M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring consensus. *Journal of Clinical Epidemiology*, 68(4), 435–441. <https://doi.org/10.1016/j.jclinepi.2014.11.021>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles, London, New Delhi: Sage.
- Geisinger, K.F. (2003). Testing and assessment in cross-cultural psychology. In J.R. Graham & J.A. Naglieri (Eds.), *Handbook of psychology* (2nd ed., pp. 95–117). Washington: John Wiley & Sons.
- Glen, S. (2014). *Cohen's kappa statistic*. Retrieved from StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/cohen-kappa-statistic/>
- Graneheim, U.H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education*, 24(2), 105–112. <https://doi.org/10.1016/j.nedt.2003.10.001>
- Hambleton, R.K. (2011). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17(3), 164–172. <https://doi.org/10.1027//1015-5759.17.3.164>
- Hernández, A., Hidalgo, M.D., Hambleton, R.K., & Gomez-Benito, J. (2020). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(2), 390–398.
- International Test Commission (ITC). (2016). The international test commission guidelines on the security of tests, examinations, and other assessments: International test commission (ITC). *International Journal of Testing*, 16(3), 181–204. <https://doi.org/10.1080/15305058.2015.1111221>
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D., Scheel, A.M., & Isager, P.M. (2018). Equivalence testing for psychological research: A tutorial. *Advances Methods and Practices in Psychological Science*, 1(2), 259–269. <https://doi.org/10.1177/2515245918770963>
- Mahmood, D., & Jacobo, H. (2019). Grading for growth: Using sliding scale rubrics to motivate struggling learners. *The Interdisciplinary Journal of Problem-Based Learning*, 13(2). <https://doi.org/10.7771/1541-5015.1844>
- Mohamed, S.A. (2013). The development of a school readiness screening instrument for grade 00 (pre-grade r) learners. Doctoral dissertation. Bloemfontein: University of the Free State.
- Munnik, E., & Smith, M.R. (2019). Methodological rigour and coherence in the construction of instruments: The emotional social screening tool for school readiness. *African Journal of Psychological Assessment*, 1, a2. <https://doi.org/10.4102/ajopa.v1i0.2>
- Munnik, E., Wagener, E., & Smith, M. (2021). Validation of the emotional social screening tool for school readiness. *African Journal of Psychological Assessment*, 3, a42. <https://doi.org/10.4102/ajopa.v3i0.42>
- Odero, E.O. (2017). Problems of finding linguistic equivalence when translating & interpreting for special purposes. *International Journal of Academic Research in Business and Social Sciences*, 7(7), 402–414. <https://doi.org/10.6007/IJARBS/v7-17/3110>
- Omar, Y.Z. (2012). The challenges of denotative and connotative meaning for second-language learners. *ETC: A Review of General Semantics*, 69(3), 324–351.
- Peters, G.J. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16(2), 59–69.
- Rawoot, I., & Florence, M.A. (2017). Equivalence and bias in the South African substance use contextual risk instrument. *Psychological Report*, 120(1), 158–178. <https://doi.org/10.1177/0033294116685865>
- Sousa, V.D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear user friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>

Appendices starts on the next page →

## Appendix 1: Quality of Translation and Linguistic Equivalence Checklist (revised)

### Section 1: Translation

#### 1.1. Experience of translators

- What percentage of the translators possessed formal qualifications translation or editing?  
None (0)      Less than 50% (1)      50% – 80% (2)      More than 80% (3)
- What percentage of the translators possessed formal qualifications in language studies?  
None (0)      Less than 50% (1)      50% – 80% (2)      More than 80% (3)

The next two items are assigned a weighted score.

#### Weighted score:

- The scores for this item will be expressed as a score out of 6
- Enter the score into the following formula
- Numerator/denominator multiplied by x/6
- Numerator: Add the scores assigned to all the translators
- Denominator: Multiply the number of translators by 3 (the maximum score possible per translator)
- x refers to the score obtained for the item

Example:

If four translators are assigned scores of 3, 3, 2 and 2, respectively, the weighted score is calculated as follows:

Numerator: Sum of the scores assigned = 10 (3 + 3 + 2 + 2)  
Denominator: 4 (number of translators) × 3 (maximum score) = 12

10/12 multiplied by x/6

$$10 \times 6 = 12x$$

$$60 = 12x$$

$$x = 5$$

NB: If the weighted score has a decimal, please round up scores with a decimal equal to and greater than 0.5 and round down scores with a decimal smaller than 0.5.

- How much experience do forward translators have in the source and target languages?

Translator 1: None (0)    Less than 5 years (1)    5–10 years (2)    More than 10 years (3)  
Translator 2: None (0)    Less than 5 years (1)    5–10 years (2)    More than 10 years (3)

Add lines for additional translators

/6

- How much experience do backward translators have in the source and target languages?

Translator 1: None (0)    less than 5 years (1)    5–10 years (2)    more than 10 years (3)  
Translator 2: None (0)    less than 5 years (1)    5–10 years (2)    more than 10 years (3)

Add lines for additional translators

/6

**Subscore 1: /18**

#### 1.2. Process of translation

- Was the forward translation conducted by at least two independent translators?  
No (0)      Yes (1)
- Were the resultant translations compared with each other?  
No (0)      Yes (1)
- How were discrepancies resolved?  
Not resolved (0)    Executive decision (1)    Discussion or consensus (2)
- Was there any external auditing conducted to confirm resolutions?  
No (0)      Yes (1)
- Was an integrated version produced in the target language?  
No (0)      Yes (1)
- Was back translation conducted?  
No (0)      Yes (1)
- Was the back translation conducted by at least two independent translators?  
No (0)      Yes (1)
- Was the integrated version produced from the forward translation used for back translation?  
No (0)      Yes (1)
- Was the back translation conducted by at least two independent translators?  
No (0)      Yes (1)
- Were the resultant translations compared with each other?  
No (0)      Yes (1)
- How were discrepancies resolved?  
Not resolved (0)    Executive decision (1)    Consensus (2)
- Was there any external auditing conducted to confirm resolutions?  
No (0)      Yes (1)

**Subscore 2: /14**  
**Section Score: /32**

**Section 2: Linguistic equivalence****2.1. Comparison between original (source document) and draft in target language**

- Was there a process to compare the original (source document) and draft in the target language?  
No (0)                      Yes (1)
- Was there a process to evaluate whether the content of items in the original version was accurately captured in the translated version?  
No (0)                      Yes (1)
- How were discrepancies about the content of items resolved?  
Not resolved (0)    Executive decision (1)    Consensus (2)
- Was there a process to evaluate whether the meanings of items in the original version were accurately captured in the translated version?  
No (0)                      Yes (1)
- How were concerns about differences in the meaning resolved?  
Not resolved (0)    Executive decision (1)    Consensus (2)
- Was there any external auditing conducted to confirm resolutions?  
No (0)                      Yes (1)
- Was there a process to compare whether the original version (source document) and translated version (target language) were equivalent in terms of:
  - Wording (vocabulary or word choice)    No (0)                      Yes (1)
  - Content    No (0)                      Yes (1)
  - Complexity                                        No (0)                      Yes (1)
  - Sentence structure                              No (0)                      Yes (1)
  - Meaning    No (0)                      Yes (1)

**Subscore 1: /13****2.2. Comparison between translated version (target language) and back translations**

- Was there a process to compare the translated version (target language) and back translation(s)?  
No (0)                      Yes (1)
- Was there a process to evaluate whether the content of items in the translated version were accurately captured in the back translations?  
No (0)                      Yes (1)
- How were concerns about the difference in the content of items in the back translations resolved?  
Not resolved (0)    Executive decision (1)    Discussion or consensus (2)
- Was there a process to evaluate whether the meanings of items in the translated version were accurately captured in the back translations?  
No (0)                      Yes (1)
- How were concerns about differences in the meaning of items in the back translations resolved?  
Not resolved (0)    Executive decision (1)    Discussion or consensus (2)
- Was there any external auditing conducted to confirm resolutions?  
No (0)                      Yes (1)
- Was there a documented process to establish whether the translated version and the back translations were equivalent in terms of:
  - Wording (vocabulary or word choice)    No (0)                      Yes (1)
  - Content    No (0)                      Yes (1)
  - Complexity                                        No (0)                      Yes (1)
  - Sentence structure                              No (0)                      Yes (1)
  - Meaning    No (0)                      Yes (1)

**Subscore 2: /13****2.3. Comparison between original version (source) and back translations**

- Was there a process to compare the original version (source document) and back translations?  
No (0)                      Yes (1)
- Was the content of items in the original version accurately captured in the back translations?  
No (0)                      Yes (1)
- How were concerns about differences in the content resolved?  
Not resolved (0)    Executive decision (1)    Discussion or consensus (2)
- Were the meanings of items in the original accurately captured in the integrated version?  
No (0)                      Yes (1)
- How were concerns about differences about the meaning resolved?  
Not resolved (0)    Executive decision (1)    Discussion or consensus (2)
- Was there any external auditing conducted to confirm resolutions?  
No (0)                      Yes (1)
- Was there a documented process to determine whether the original and back translations were equivalent in terms of:
  - Wording (Vocabulary or word choice)    No (0)                      Yes (1)
  - Content    No (0)                      Yes (1)
  - Complexity                                        No (0)                      Yes (1)
  - Sentence structure                              No (0)                      Yes (1)
  - Meaning    No (0)                      Yes (1)

**Subscore 3: /13  
Section score: /39**

## Appendix 2: Quality of Translation and Linguistic Equivalence Checklist: Interpretation Matrix

Rating	Poor	Good	Excellent
<b>Section 1: Translation</b>			
Score (SS1 + SS2) (___/32)	< 50% < 16	50% – 79% 16–24	80% – 100% 25+
<b>Quality level</b>	<b>Low level of compliance with ITC guidelines</b>	<b>Basic compliance with ITC guidelines</b>	<b>High level of compliance with ITC guidelines</b>
<b>Action</b>	<b>Redo as per recommended guidelines</b>	<b>Identify and revise items where concerns have been raised</b>	<b>Proceed to establish equivalence</b>
<b>Section 2: Linguistic equivalence</b>			
Score	< 50%	50% – 79%	80% – 100%
Subscores (___/13)	< 6 < 19	6–10 19–30	11–13 31–39
<b>Composite score</b> (SS1 + SS2 + SS3) (___/39)			
<b>Quality assurance</b>	<b>Low level of equivalence</b>	<b>Basic equivalence</b>	<b>High level of equivalence</b>
<b>Action</b>	<b>Redo as per recommended guidelines</b>	<b>Identify and revise items where concerns have been raised.</b>	<b>Conclude high level of linguistic was achieved</b>

## Appendix 3: Quality of Translation and Linguistic Equivalence Checklist template

Translation		
Criterion	Description	Response
<i>Formal qualifications translation or editing</i>	Indicate the number of translators. How many translators have formal qualifications in translation or editing? Express as a percentage.	
<i>Formal qualifications in language studies</i>	Indicate the number of translators. How many translators have formal qualifications in language studies? Express as a percentage.	
<i>Cumulative experience in the source and target languages</i>	Indicate number of years of experience in translation for each translator. <ul style="list-style-type: none"> <li>• Translator 1</li> <li>• Translator 2</li> <li>• Translator X (add lines for each additional translator)</li> </ul>	
<i>Forward translation</i>	How many translators were involved with forward translation? Did the translators conduct their translations independently? Were the resultant translations compared to each other? Describe how discrepancies between the translations were resolved. Were the resolutions presented to an external person for confirmation and verification? Was a single translated version produced from the forward translations after comparison?	
<i>Back translation</i>	Was back translation conducted? How many translators were involved with back translation? Did the translators conduct their translations independently? What was used as the basis for the back translation? Were the resultant translations compared with each other? Describe how discrepancies between the translations were resolved. Were the resolutions presented to an external person for confirmation and verification?	
<i>Additional notes</i>	Report anything that you considered during translation that was not assessed here.	

## Appendix 4: Quality of Translation and Linguistic Equivalence Checklist: Reviewer response form

Translation		
Criterion	Rating	Score
<b>Experience of translators</b>		
What percentage of the translators possessed formal qualifications translation or editing?	None (0) Less than 50% (1) 50% – 80% (2) More than 80% (3)	
What percentage of the translators possessed formal qualifications in language studies?	None (0) Less than 50% (1) 50% – 80% (2) More than 80% (3)	
The next two items are assigned a weighted score.	Example: If four translators are assigned scores of 3, 3, 2 and 2, respectively, the weighted score is calculated as follows: Numerator: sum of the scores assigned = 10 (3 + 3 + 2 + 2) Denominator: 4 (number of translators) × 3 (maximum score) = 12 10/12 multiplied by x/6 10 × 6 = 12x 60 = 12x x = 5	NB: If the weighted score has a decimal, please round up scores with a decimal equal to and greater than 0.5 and round down scores with a decimal smaller than .5.
Weighted Score: • The scores for this item will be expressed as a score out of 6 • Enter the score into the following formula • Numerator/denominator multiplied by x/6 • Numerator: Add the scores assigned to all the translators • Denominator: Multiply the number of translators by 3 (the maximum score possible per translator) • x refers to the score obtained for the item		
How much experience do forward translators have in the source and target languages?	Use for every translator: None (0) Less than 5 years (1) 5–10 years (2) More than 10 years (3)	T1 - T2 - Item score /6
How much experience do backward translators have in the source and target languages?	Use for every translator: None (0) Less than 5 years (1) 5–10 years (2) More than 10 years (3)	T1 - T2 - Item score /6
<b>Subscore 1</b>		
<b>Process of translation</b>		
Was the forward translation conducted by at least two independent translators?	No (0) Yes (1)	
Were the resultant translations compared to each other?	No (0) Yes (1)	
How were discrepancies resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Was there any external auditing conducted to confirm resolutions?	No (0) Yes (1)	
Was an integrated version produced in the target language?	No (0) Yes (1)	
Was back translation conducted?	No (0) Yes (1)	
Was the back translation conducted by at least two independent translators?	No (0) Yes (1)	
Was the integrated version produced from the forwards translation used as a basis for the back translation?	No (0) Yes (1)	
Were the resultant translations compared to each other?	No (0) Yes (1)	
How were discrepancies resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Was there any external auditing conducted to confirm resolutions?	No (0) Yes (1)	
<b>Subscore 2</b>		
<b>Section Score</b>		
<b>Quality description</b>		
<b>Action</b>		
<b>Linguistic equivalence</b>		
Criterion	Rating	Comment
<b>Comparison between original (source document) and draft in target language</b>		
Was there a process to compare the original (source document) and draft in the target language?	No (0) Yes (1)	
Was there a process to evaluate whether the content of items in the original version were accurately captured in the translated version?	No (0) Yes (1)	
How were concerns about differences in the content resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	

Appendix 4 continues on next page →

**Linguistic equivalence**

Criterion	Rating	Comment
Was there a process to evaluate whether the meanings of items in the original version were accurately captured in the translated version?	No (0) Yes (1)	
How were concerns about differences in the meaning resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Was there any external auditing conducted to confirm resolutions?	No (0) Yes (1)	
Was there a process to compare the original version (source document) and translated version (target language) were equivalent in terms of		
Wording (vocabulary or word choice)	No (0) Yes (1)	
Content	No (0) Yes (1)	
Complexity	No (0) Yes (1)	
Sentence structure	No (0) Yes (1)	
Meaning	No (0) Yes (1)	

**Subscore 1****Comparison between translated version (target language) and back translations**

Was there a process to compare the translated version (target language) and back translation(s)?	No (0) Yes (1)	
How were concerns about the difference in the content of items in the back translations resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
How were concerns about differences in the meaning of items in the back translations resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Was there any external auditing conducted to confirm resolutions?	No (0) Yes (1)	
Was there a documented process to establish whether the translated version and the back translations were equivalent in terms of		
Wording (vocabulary or word choice)	No (0) Yes (1)	
Content	No (0) Yes (1)	
Complexity	No (0) Yes (1)	
Sentence structure	No (0) Yes (1)	
Meaning	No (0) Yes (1)	

**Subscore 2****Comparison between original version (source) and back translations**

Was there a process to compare the original version (source document) and back translations?	No (0) Yes (1)	
Was the content of items in the original accurately captured in the integrated version?	No (0) Yes (1)	
How were concerns about differences in the content resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Were the meanings of items in the original accurately captured in the integrated version?	No (0) Yes (1)	
How were concerns about differences about the meaning resolved?	Not resolved (0) Executive decision (1) Discussion or consensus (2)	
Was there any external auditing conducted to confirm resolutions?	No (0) Yes (1)	
Was there a documented process to establish whether the translated version and the back translations were equivalent in terms of		
Wording (Vocabulary or word choice)	No (0) Yes (1)	
Content	No (0) Yes (1)	
Complexity	No (0) Yes (1)	
Sentence structure	No (0) Yes (1)	
Meaning	No (0) Yes (1)	

**Subscore 3****Section score****Quality description****Action**

## Appendix 5: Reviewer's comments

Translation			
Criterion	Rater 1	Rater 2	Solved
<b>Experience of raters</b>			
How much cumulative experience do the translators have in the source and target languages?	The aggregation here might not be meaningful.	The total might skew the picture and not differentiate between different translators.	✓
<b>Linguistic equivalence</b>			
<b>Comparison between translated version (target language) and back translations</b>			
Were there any items where there were concerns about the content of items in the back translations?			✓
If yes, how was it resolved?	The response options are not optimally described, which makes the scoring difficult.	The description provided on the template did not explicitly state that they reached consensus.	
Were there any items where there were concerns about the meaning of items in the back translations?			
If yes, how was it resolved?	I assumed that consensus would include discussion and therefore awarded the higher mark.	This was unclear and a clearer description of consensus and discussion was needed.	✓
Comparison between original version (source) and back translations			
If yes, how was it resolved?	See earlier comments about the scoring options.	The same concerns as on previous items using discussion and consensus.	✓
<b>Qualitative comments</b>			
<b>Reviewer 1</b>			
The checklist was clear and easy to use. The one concern was that the lack of description of response options related to resolution of differences impacted scoring. The subsection and section scores did not add up. The reviewer form and template made evaluation easy and efficient. Check the totals for subsections and sections. It does not always add up.			
<b>Reviewer 2</b>			
The inclusion of the template was useful so that the researchers can capture their process for evaluation. The reviewer form was very useful as it had a high level of alignment, which made scoring easier.			
Consider revising the answer options on the questions related to resolving differences to avoid subjective interpretations. The descriptions of quality are useful. The scores for the ranges must be revised to ensure that it is accurate.			