AOSIS

Cross-cultural adaptation and validation of the STarT back screening tool in isiZulu

CrossMark
click for updates

Authors:
Peta-Ann Schmidt[1]
Vaneshveri Naidoo[1]

Affiliations:
[1]Department of Physiotherapy, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding author:
Peta-Ann Schmidt,
paschmidt@mweb.co.za

Background: Non-specific low back pain (NSLBP) is one of the most prevalent conditions in the world. Identifying patients at risk for developing chronic NSLBP is key to effective treatment. The STarT back screening tool is a validated, prognostic screening tool identifying subgroups of NSLBP patients, and the risk factors associated with each subgroup, guiding treatment in the primary care of NSLBP.

Objectives: To translate the English version of the STarT back screening tool into isiZulu and determine the content validity and reliability of the translated tool.

Method: Translation was completed in four phases - forward translation and synthesis, backward translation and expert review. Validation included expert review for content validity and testing of the translated tool on 30 patients, determining test-retest reliability, internal consistency and usability.

Results: Minor linguistic differences were addressed during the translation phase. Item content validity was excellent for relevance (1.00), satisfactory (0.94) for clarity, simplicity and ambiguity, with scale-content validity acceptable (0.955). Spearman's correlation coefficient for test-retest reliability was acceptable (0.73). Cronbach's alpha for internal consistency for the total score for test 1 and test 2 was 0.68 and 0.77, and for the psychosocial scale 0.62 and 0.77 respectively. Overall, 33% found the tool very easy to understand and 40% found it very easy to complete.

Conclusion: The isiZulu STarT back screening tool showed excellent content validity, acceptable reliability and acceptable internal consistency.

Clinical implications: Use of the isiZulu tool in local clinics and private practices can improve clinical decision-making and treatment outcomes for isiZulu-speaking patients with NSLBP.

Keywords: STarT back screening tool; isiZulu; NSLBP; validity; reliability; translation.

Introduction

Non-specific low back pain (NSLBP) is one of the most prevalent conditions in the world (Bruyere et al. 2012) and was the leading cause of disability worldwide in 2015 (Global Burden of Disease (GBD) Study 2015). The prevalence of NSLBP in African populations has been shown to be similar to the rest of the world (Louw et al. 2007). Many patients experience short episodes of acute NSLBP that resolve quickly, but others go on to develop chronic NSLBP, which can persist for up to 1 year after the initial onset (Hill et al. 2008). Approximately 23% of patients experience ongoing symptoms, of which 11% – 12% continue to experience long-term disability (Airaksinen et al. 2006). Chronic NSLBP thus places a significant socio-economic burden on society (Melloh et al. 2009), making diagnosis of patients at risk of developing chronic NSLBP very important.

Predicting the development of chronicity in NSLBP through the early identification of risk factors (Van Tulder et al. 2006) has been shown to lead to more effective treatment interventions (Melloh et al. 2009). Emphasis on psychosocial factors as risk factors for developing chronic NSLBP occurs in most low back pain clinical guidelines (Koes et al. 2010), and identification of the presence of such factors early in the course of NSLBP has been shown to improve outcomes of intervention (Hayden et al. 2010). Hill et al. (2008) developed and validated a back pain prognostic screening tool, the STarT (subgroups for targeted treatment) Back Screening Tool, which could be used to identify subgroups of patients with NSLBP and their associated risk factors (Keele University 2017), grouping these patients into low-, medium- and high-risk groups. A randomised controlled trial by Hill et al. (2011) compared a stratified care approach based on this back pain prognostic screening tool with current best practices and showed a significant improvement in patient outcomes and substantial economic benefits.

The STarT back screening tool is a self-administered questionnaire, and understanding the written language of this standardised tool is integral to its usage and interpretation. Understanding a language, written or spoken, involves the need to integrate new information with what is already available (Van Dyke 2016). Translation therefore gives different cultures the opportunity to access a huge amount of new information available in the English world, and in return allows the English world an understanding of other cultures. The STarT back screening tool was originally developed and validated in English and has been translated into 35 languages worldwide (Keele University 2017), and there is only one translated version of this tool available in any of the other official languages of South Africa, namely Afrikaans (The 11 languages of South Africa 2019).

Most South Africans, particularly in the African population, are bilingual and often use more than one language at home, with English usually considered a second or third language, and if a person is not able to use the language most familiar to them, they are unlikely to perform at their best, and their self-confidence will be undermined (The Importance of languages 2017).

Therefore, understanding the STarT back screening tool in their first language allows patients with NSLBP to use the tool more effectively, and it would assist practitioners to accurately group patients into low-, medium- and high-risk groups, which, in turn, would direct treatment interventions. Subsequently, validating the translated tool would determine how accurately it measures what it intends to measure. The objectives of this study were to cross-culturally adapt the STarT back screening tool into isiZulu and determine the content validity and reliability of the translated tool in an isiZulu population of NSLBP patients.

Research methods and design

This study consisted of two phases – a cross-cultural adaptation phase, which included forward translation, translation synthesis, back translation and expert committee review, and a validity phase, which included testing of content validity, test–retest reliability, internal consistency and usability.

Participants of this study were purposively sampled on the basis of the guidelines established by Beaton et al. (2000) for the process of cross-cultural adaptation of self-reported measures. The participants in the cross-cultural adaptation phase of this study included two forward translators known to the researcher, two backward translators from Wits Translate (Wits Language Schools' professional translation agency), and three healthcare professionals from the Department of Physiotherapy, who all made up the expert review committee.

Inclusion criteria for forward translation were a medically trained isiZulu first-language speaker or a university educated, non-medically trained isiZulu first-language speaker. For backward translation, services of two second-language English-speaking individuals, proficient in isiZulu and non-medically trained were utilised. For expert review, services of one healthcare professional proficient in isiZulu and English, two healthcare professionals proficient in English, with translation experience and the aforementioned forward and back translators were utilised.

The two forward translators translated the original English STarT back screening tool into isiZulu, developing translation 1 (T1) and translation 2 (T2). T1 and T2 were compared and discrepancies discussed until consensus on a common translation was reached. Synthesis of the two translated screening tools (T1 and T2) was conducted, and a new screening tool (T12) was produced. The two backward translators translated the T12 screening tool back into English, producing translations BT1 and BT2. The expert committee reviewed all the screening tools (T1, T2, T12, BT1 and BT2), discussed discrepancies and consolidated all the screening tools to develop a preliminary translated isiZulu STarT back screening tool.

The validation phase used a sample of convenience. For content validity, two forward translators, one backward translator and one healthcare professional from the expert committee were included. For test–retest reliability, internal consistency and usability, isiZulu South African men and women aged 18 years and older with acute or chronic NSLBP, with or without associated leg pain, willing to participate in this study, and able to return 2 weeks later for re-testing were included. They were recruited from Erasmus Physiotherapists Inc., Witbank General Hospital, Siphosensimbi Community Health Clinic and Lynnville Polyclinic (Mpumalanga Department of Health) in Witbank/eMalahleni. Non-university-educated, non-Zulu-speaking and reading persons were excluded when testing for content validity. Patients with potentially serious pathology (cauda equina, cognitive, neurological or rheumatologic disease, fracture of the spine and malignancy), current specific diagnoses such as disc herniation or spondylolisthesis, spinal surgery in the last 6 months, pregnancy, and those who do not read or speak isiZulu were excluded when testing test–retest reliability, internal consistency and usability.

Polit and Beck (2006) and Yaghmaie (2003) recommended a panel of four experts as the minimum number of participants required to determine item content validity (I-CVI) and scale-content validity (S-CVI). The minimum total sample size to determine test–retest reliability, internal consistency and usability was calculated to be 27 patients based on an intra-class correlation coefficient (ICC) of 0.9 and a maximum width of 0.23 ($N = [16p\,(1-\pi)]/w^2$, where p is the expected ICC of 0.9 and w is the width of 0.23). The expected ICC of 0.9 and the maximum width of 0.23 were based on translation and validation studies conducted in Xhosa, Afrikaans and English populations by Morris et al. (2012).

The preliminary translated isiZulu STarT back screening tool was reviewed for I-CVI by the four experts and then tested on the intended patient population as recommended by Beaton et al. (2000).

Two weeks later a re-test of the preliminary translated isiZulu version on the same patient population was performed as conducted by Hill et al. (2008) in the validation of the English STarT back screening tool. Transport costs for each participant were reimbursed by the first author. A post-retest questionnaire was administered at this point to determine the feasibility (utility, efficiency, learnability and satisfaction [Nielsen 2012]) of the isiZulu STarT back screening tool (Glover & Albers 2007), where each patient reported on the ease of comprehension of the translated tool, the ease of completing the tool, the time needed to complete the tool (5, 10, 15 and 20 min) and any suggestions to improve the clarity of the tool. The participants were sent reminders via SMS to attend follow-up visits at the same place where they were recruited. There was no control for confounding factors during the 2 weeks until follow-up. After all testing and re-testing were completed, the pre-final version of the translated isiZulu STarT back screening tool was submitted to the original developers of the tool for approval of study methodology and outcome. Once the original developers of the STarT back screening tool were satisfied with the translation and validation process, the translated isiZulu tool was made available on the following website: https://startback.hfac.keele.ac.uk/training/resources/start-back-translations/.

The content validity was determined by calculating I-CVI and S-CVI, measuring the representativeness and comprehensiveness of the items in a scale (Yaghmaie 2003) as well as the degree to which the instrument measures a specific construct (Bolarinwa 2015). A four-point ordinal scale was used to rate each question/item based on relevance, clarity, simplicity and ambiguity (Yaghmaie 2003), which included the following: (1) not relevant (not clear/not simple/doubtful); (2) question needs some revision; (3) relevant (clear/simple/no doubt but needs minor revision); and (4) very relevant (clear/simple/meaning is clear). The item content validity was calculated as the number of experts giving a rating of 3 or 4 divided by the total number of experts for each question/item. The item content validity should be 1.00 when there are five or fewer experts (Polit & Beck 2006). This was carried out to ensure that the translated version of the STarT back screening tool was understood in the local context and language. Scale-content validity was subsequently calculated on the basis of these results by adding the I-CVI for each aspect (relevance, clarity, simplicity and ambiguity) and dividing by the total number of aspects to arrive at the average (I-CVI for relevance + I-CVI for clarity + I-CVI for simplicity + I-CVI for ambiguity/4). The S-CVI is the proportion of the total number of items that are content-valid, or the proportion of items that received a score of 3 or 4 by the experts (Polit & Beck 2006), and is thus the average number of items rated as 3 or 4 across the questions/items for relevance, clarity, simplicity and ambiguity (Polit & Beck 2006). The S-CVI of 0.80 or higher is considered acceptable (Polit & Beck 2006).

The test–retest reliability was measured using Spearman's correlation coefficient. Correlation coefficients higher than 0.8 demonstrate excellent reliability, coefficients between 0.6 and 0.8 demonstrate good reliability, and coefficients below 0.6 show poor reliability. The Internal Consistency Index of reliability/homogeneity is the degree to which the questions or items in the tool measure the same thing (Bolarinwa 2015). The higher the Internal Consistency Index or reliability value, the more reliable the measure. The Internal Consistency Index of this tool was evaluated by calculating Cronbach's alpha coefficients for the total score (questions 1–9) and for the psychosocial subscale (questions 5, 6, 7, 8 and 9). An alpha index equal to or more than 0.90 is considered excellent (Bolarinwa 2015), equal to or more than 0.80 is considered good, equal to or more than 0.70 is considered acceptable, equal to or more than 0.60 is considered questionable, whilst less than 0.60 shows poor internal consistency.

Descriptive statistics analysed the data regarding socio-demographic information and previous history of low back pain and the information pertaining to the usability of the tool. Information about dropouts and missing data was also recorded.

Ethical consideration

Ethical approval to conduct the study was obtained from the Health Research Ethics Committee (HREC) of the University of the Witwatersrand (Ethical Clearance Number: M170717), and permission to translate and validate the STarT back screening tool into isiZulu was obtained from the original authors at the Keele University in North Staffordshire, UK.

Results
Phase 1: Translation

Minor linguistic differences emerged during the forward and back translation phases, which were addressed in consultation with the expert committee. The word 'sometime' in questions 1 and 2 of the English tool was translated to mean 'other time' in isiZulu, with no word meaning exactly the same in isiZulu. The words 'more slowly' in question 4 were translated to mean 'long time' in isiZulu, as again no isiZulu word was found for 'slowly'. In question 5, the words 'physically active' were initially translated to mean 'mostly exercise' with a bit of other activity, but the expert committee agreed that it relates to being physically active during activities of daily living (ADLs) and not specific exercises. The translation was thus changed to mean 'general activity', which better aligns with the original text. The translators agreed that the term 'worrying thoughts' in question 6 could be translated to mean worry or stressful. It was decided that *engikhatazayo* which means 'bothering me' was the closest translation, as worry and stressful are understood as the same concept in isiZulu. The forward and back translators found the word 'terrible' in question 7 difficult to translate as no specific word exists in isiZulu. The expert committee agreed to *ngizwasengathi*, which means 'I feel like' and better covers the original meaning of the word. Difficulty was found in translating 'in general' and 'enjoy' in question 8, as no direct translated words exist in isiZulu. The expert

committee revised the full sentence to convey the meaning of both these terms – *Angizange ngizithokozele izinto engivame ukuzijabulisa ngazo njengenjwayelo*, which translates as 'I did not enjoy the things I usually enjoy'. The word 'bothersome' in question 9 is not a word often used in Africa, and does not exist in isiZulu, so the expert committee agreed that the most common translation *kukuhluphe kangakanani* will be used, meaning 'to bother or trouble'.

The above changes were made, and the preliminary final translation was resubmitted to the expert committee for editing. No further changes were required, and this preliminary final translation was used in the validation process.

Phase 2: Validation

A summary of the demographic characteristics of the sample population is given in Table 1. A total of 37 patients was recruited to participate in our study, with seven patients not returning for the second round of testing. Two questionnaires were declared invalid because of missing data. Overall, 76% of the sample population consisted of women with 40% aged between 50 and 59 years. Educationally, 53% were high school educated, 60% were unemployed, and 50% had no previous history of back pain. Finally, 60% of the sample population was allocated to the high-risk group after tests 1 and 2.

TABLE 1: Demographic characteristics (*n* = 30).

Variable	Category	Frequency (*n*)	%
Gender	Male	7	23
	Female	23	76
Age	21–29 years	4	13
	30–39 years	2	6
	40–49 years	8	26
	50–59 years	12	40
	60 years and older	4	13
Level of education	Primary school	10	33
	High school	16	53
	University	4	13
Occupation	Sedentary	3	10
	Non-sedentary	9	30
	Unemployed	18	60
Hours worked	40 h per week	7	23
	Less than 40 h per week	4	13
	Retired	3	10
Previous history of back pain	None	15	50
	Falls	5	16
	Physical violence	1	3
	Motor vehicle accidents (MVAs)	2	6
	Back and neck injury	2	6
Risk groups Test 1	Low risk	6	20
	Medium risk	6	20
	High risk	18	60
Test 2	Low risk	6	20
	Medium risk	4	13
	High risk	20	60

TABLE 2: Test–retest reliability.

Sample size (*n*)	Spearman's correlation coefficient (ρ)	Significance level (*p*)	95% confidence interval (95% CI)
30	0.73	0.0001	0.381–0.820

Content validity

The content validity was measured in two components, I-CVI and S-CVI. The item content validity for relevance was excellent (means I-CVI = 1.00) and was satisfactory for clarity, simplicity and ambiguity (means I-CVI = 0.94). The S-CVI was calculated by adding the I-CVI for each category (relevance, clarity, simplicity and ambiguity) and dividing by the total number of categories to reach the average. The S-CVI for the final translated isiZulu STarT back screening tool was 0.955, agreeing with Polit and Beck's (2006) recommendation of an S-CVI being considered acceptable with a value of 0.80 or higher.

Test–retest reliability

The test–retest reliability was determined by calculating Spearman's correlation coefficient (ρ) between test 1 and test 2, resulting in ρ = 0.73, which demonstrates *acceptable* test–retest reliability (Glen 2018) (Table 2).

Internal consistency

The internal consistency was evaluated by calculating Cronbach's alpha coefficients (item by item) for questions 1–9. An alpha index of more than 0.70 is considered to be acceptable internal consistency (Bolarinwa 2015).

Cronbach's alpha was calculated for the total score (questions 1–9) as 0.68, which indicates less than acceptable internal consistency, and for the psychosocial subscale (questions 5, 6, 7, 8 and 9) as 0.68, which also showed less than acceptable internal consistency.

Usability

Figure 1 demonstrates usability of the developed questionnaire: 33% found the translated tool very easy to understand, whilst 3% found it very difficult to understand.

Figure 2 demonstrates the ease of completion of the developed questionnaire: 40% found the translated tool very easy to complete, whilst 10% found it moderately difficult to complete.

Figure 3 demonstrates the time taken to complete the developed questionnaire: 20% took 20 min to complete the
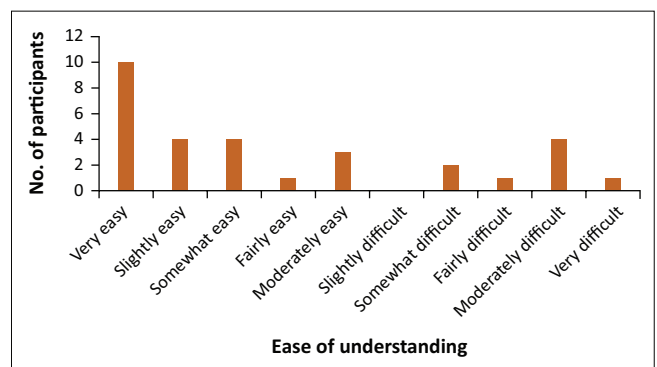

FIGURE 1: Ease of understanding the translated tool (*n* = 30).

translated isiZulu tool, 13% completed it in 15 min, 44% took 10 min and 23% completed it in 5 min.

Table 3 demonstrates suggestions from the participants: 27% of the participants requested assistance with the explanation of the questions.

Discussion
Phase 1: Translation

The translation procedure applied in this study followed the guidelines set out by Beaton et al. (2000). The English STarT back screening tool and the final isiZulu STarT back screening tool are provided in Figures 4 and 5, respectively.

Finding equivalent words without altering the meaning of the original sentence was the major problem experienced

FIGURE 2: Ease of completion of the translated tool (*n* = 30).

1. 5 minutes (23%)
2. 10 minutes (44%)
3. 15 minutes (13%)
4. 20 minutes (20%)

FIGURE 3: Time to complete the translated tool (*n* = 30).

TABLE 3: Participants' suggestions (*n* = 30).

Suggestions	Frequency (*n*)	Percentage (%)
None	19	63
Include more questions about lifting, bending and standing	1	3
Prefer tool in English	1	3
Need assistance with explanation of questions	8	27
Title of tool needs to be translated	1	3

during the translation phase and was resolved through discussions within the expert committee.

Firstly, the closest possible translation for a specific word was used; secondly, the closest possible translated meaning for a phrase was used; and thirdly, sentence structure was adjusted to convey the closest possible meaning to the English version. Some words had no equivalent in isiZulu and were translated to the closest possible meaning, such as 'sometimes' in questions 1 and 2, 'worrying' in question 6 (similar difficulty was faced in the French study; Bruyere et al. 2012), 'terrible' in question 7 (similar difficulty was faced in the Finnish study; Piironen et al. 2016), 'enjoy' in question 8 and 'bothersome' in question 9 (similar difficulty was faced in the Finnish, Piironen et al. 2016, and the Danish, Morso et al. 2011 studies). Certain phrases such as 'physically active' in question 5 posed difficulties for both forward and back translators, as well as 'in general' in question 8. Similar difficulties were experienced in the German (Aebischer et al. 2015), Swedish (Betten et al. 2015), Italian (Maggiani & Abenavoli 2019) and Finnish (Piironen et al. 2016) translation studies. Ultimately the expert committees in all these studies decided to keep the wording as close to the original English version as possible so as not to lose the meaning entirely. All translation challenges and changes to the isiZulu version were addressed and approved by the expert committee before testing the validity, reliability and usability of the translated tool.
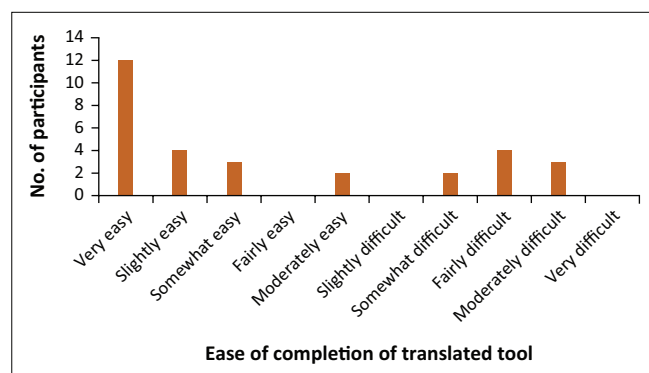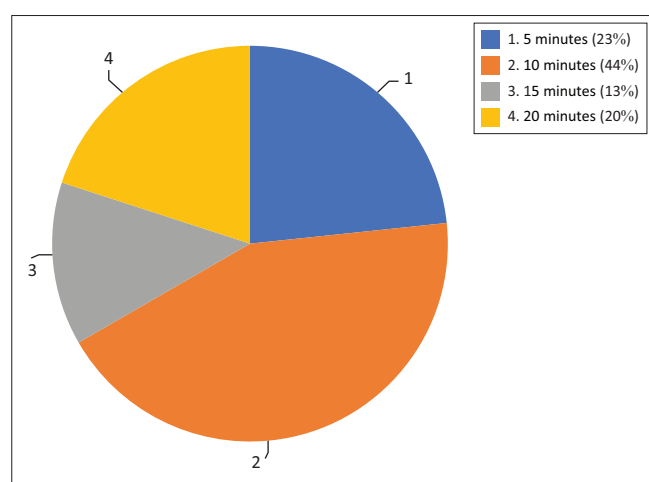
Phase 2: Validation

South Africa has one of the most diverse populations on the African continent, with 11 official recognised languages.

FIGURE 4: English STarT back screening tool.

FIGURE 5: isiZulu STarT back screening tool.

Almost one-quarter (23%) of the population speaks isiZulu as their home language, and 58% speak it as their second language. English is the dominant language in major cities across South Africa but is only spoken as a second language by black South Africans in most other and rural areas of the country, including Witbank/eMalahleni, Mpumalanga, where up to 24% speak isiZulu (The 11 languages of South Africa 2019).

The majority of the sample population consisted of women (77%) aged between 50 and 59 years, which is similar to the literature showing a higher prevalence of low back pain onset in women aged between 30 and 50 years (Nicholas et al. 2011; Yoshimoto et al. 2017), and which does not differ in the developed world versus the developing world (Louw et al. 2007). A number of other translation studies identified employment status of their participants (Hill et al. 2008; Karstens et al. 2015; Piironen et al. 2016; Robinson et al. 2017), highlighting that the majority of patients were employed or received disability benefits, whereas most participants (60%) in our study were unemployed and did not receive employment or disability benefits.

This is an expected difference, as our study is one of the only studies carried out in developing Africa, where resources are limited, income inequality exists and the majority of the population does not have access to disability benefits or unemployment funds (United Nations Development Program (UNDP) 2017).

Most of the sample population was high school educated (53%), which was similar to the Swedish (64%) (Betten et al. 2015), Portuguese (32%) (Raimundo et al. 2017) and

Brazilian–Portuguese (60%) (Pilz et al. 2014) studies. None of the other translation studies, including the original English study, examined the educational levels of their sample populations, perhaps suggesting that educational levels would not influence their translation process. Educational levels may not be a concern in developed countries; however, in the developing world where disparity in education exists, level of education/literacy could impact on completing the questionnaire (United Kingdom [UK] Essays 2016). The legacy of apartheid presents a significant challenge for education and health systems in South Africa. Black South Africans people were on the receiving end of education inequality, resulting in lower educational levels, a lack of educational resources and a shortage of qualified teachers (UK Essays 2016). This has led to a lack of basic information regarding health and wellness, particularly amongst those above the age of 50 years (UK Essays 2016). This is demonstrated by the majority of this study's participants (77% in the 50–59-year age category), and the age prevalence of NSLBP (30–60 years) (Maher, Underwood & Buchbinder 2017). Lower levels of education demonstrated in our study suggest that the participants do not have access to basic healthcare information as a result of education inequalities and healthcare challenges.

Most participants were stratified into the high-risk group (60%), which may suggest that the patients are not actively seeking treatment for their back pain, were not getting the correct treatment when their back pain began, or were unaware of the back care treatment options available. The possibility also exists that the staff at local primary care clinics in Mpumalanga was not educated on the correct treatment pathways for these patients, and thus did not know that these patients should be referred for physiotherapy or for interdisciplinary management.

Measuring a questionnaire's content validity measures the questionnaire's quality (Bolarinwa 2015), which is a necessary part of developing an appropriate tool. Acceptable content validity shows that the content is relevant and appropriate, particularly when adapting a specific tool culturally, as seen in our study. The Dutch translation study (Bier et al. 2017) and the original English tool (Hill et al. 2008) tested content validity based on floor and ceiling effects, contrary to our study. For floor and ceiling effects to be present, some items need to be missing in the lower and upper ends of the scale, which results in limited content validity (Terwee et al. 2007). Floor and ceiling effects were not tested in our study, as the recommended sample size (Terwee et al. 2007) to test these effects is at least 50 patients. This study comprised only 30 patients, and the sample size calculation in our study was primarily based on a similar translation study performed by Morris et al. (2012).

Other translation studies (Morris et al. 2012) tested face and content validity simultaneously, using a group of patients to report the ease of understanding and ease of completing the translation. Our study did not test face validity, as it is an even more subjective test than content validity (Bolarinwa

2015), but used a panel of experts to test content validity, and the sample population to report on usability of the isiZulu tool. Criterion-related (predictive and concurrent) validity and construct validity were tested in other translation studies (Mphahlele, Mitchell & Kamerman 2008, Sharafi et al. 2017; Tsang Chi Chung et al. 2017), but were not tested in our study because the STarT back screening tool is a screening tool and does not predict future change. Furthermore, there is a lack of single reference isiZulu translated standard tools that screens for similar constructs.

The test–retest reliability of the preliminary isiZulu STarT back screening tool showed acceptable reliability (ICC of 0.73; 95% confidence interval [CI]: 0.38–0.82) for the total score, which is similar to that of the original version for the total score (ICC of 0.73; 95% CI: 0.57–0.84) (Hill et al. 2008). The Brazilian–Portuguese (Pilz et al. 2014), Portuguese (Raimundo et al. 2017), Finnish (Piironen et al. 2016) and German (Karstens et al. 2015) studies also demonstrated acceptable reliability (ICC ranging from 0.70 to 0.78).

Internal consistency in our study showed less than acceptable results for both the total score (0.68) and the psychosocial subscale (0.68). This was less than the original English version (0.74 total score and 0.74 psychosocial subscale) (Hill et al. 2008) as well as the Brazilian–Portuguese (Pilz et al. 2000), Portuguese (Raimundo et al. 2017), Finnish (Piironen et al. 2016), French (Bruyere et al. 2012), Persian (Abedi et al. 2015) and Japanese (Matsudaira et al. 2016) versions. Difference in the results of our study could be because of the way the constructs of the original English version were translated, with many words and phrases not existing in the isiZulu language. It may also be because of the level of education of our study sample not understanding the translated concepts clearly, particularly the negative wording of some of the questions (Bolarinwa 2015). Less than acceptable internal consistency resulted in less than acceptable reliability of the tool, which suggests the need to retest the translated tool on a different sample population.

Usability, tested in 30 participants, was found to be acceptable, easy to understand and easy to complete, whereas all the other translated versions, including the original English version (Hill et al. 2008), used a smaller sample size (n = 12–30), and also found good acceptability, ease of understanding and ease of completion (Abedi et al. 2015; Aebischer et al. 2015; Betten et al. 2015; Bier et al. 2017; Bruyere et al. 2012; Gusi et al. 2011; Hill et al. 2008; Karstens et al. 2015; Luan et al. 2014; Maggiani & Abenavoli 2019; Morso et al. 2011, Piironen et al. 2016; Pilz et al. 2000; Raimundo et al. 2017). The participants in our study took longer to complete the isiZulu questionnaire than the other aforementioned translation studies, which may be because of educational levels noted in the sample population, or because of this being the first time this sample population has been exposed to any form of health questionnaire.

The strengths of our study included the use of internationally recognised guidelines, the use of first-language isiZulu speakers throughout the process, which strengthened the linguistic accuracy of the translation, as well as the use of a wide inclusion criterion for the validation phase, and patient recruitment from a variety of geographical outpatient clinics which contributed to the acceptable external validity and reliability of our study. The limitations of our study included its inability to measure criterion or construct validity because of the lack of a single reference standard isiZulu translated tool against which the isiZulu STarT back screening tool could be tested. A smaller sample size than recommended by Terwee et al. (2007) could also have limited the validation results. The isiZulu tool also did not account for the treatment effect that NSLBP patients may have received between the first and second test–retest reliability testing, which could affect the outcome of the test–retest reliability test. However, the original STarT back screening tool was not designed to gather this type of information. The isiZulu tool was also only tested in the Mpumalanga province of South Africa, where many but not the majority of isiZulu-speaking people reside.

It is necessary to conduct another study to evaluate the psychometric properties in a larger, representative sample. Translation and validation of the Orebro Musculoskeletal Pain Screening Questionnaire (OMPSQ) into isiZulu, which then could be used as a reference standard to test the isiZulu STarT back screening tool could assist in fully validating the tool. It is also necessary to test the clinical effectiveness and cost-effectiveness of the isiZulu translation as well as the original English version in a South African context by comparing this stratified approach with a non-stratified best practice approach. This would determine whether stratified care leads to a significant decrease in healthcare costs and faster recovery periods specifically in South Africa.

isiZulu is the most widely spoken language in South Africa, so the widespread use of the isiZulu STarT back screening tool could have a significant effect on the way NSLBP is treated. Therefore, making the isiZulu tool available for use by clinicians in local clinics and private practices would improve clinical decision-making and treatment outcomes for isiZulu-speaking patients with NSLBP, and possibly improve physiotherapy and inter-professional referrals as needed. Training in 'psychologically informed physiotherapy' is also needed for treatment of the high-risk group.

Conclusion

Established guidelines for the translation process were followed, and a preliminary isiZulu STarT back screening tool was produced, which showed good linguistic accuracy. This preliminary isiZulu tool was tested on a sample population of isiZulu-speaking patients presenting with NSLBP, and they reported good understanding and ease of completing the tool. The translated tool showed good content validity, acceptable test–retest reliability and less than acceptable internal consistency. Therefore, the preliminary isiZulu STarT back screening tool could be used in an isiZulu population presenting with NSLBP.

Acknowledgements

Data availability statement

The authors confirm that the data supporting the findings of this study are available within this article and/or in its supplementary materials.

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

Abedi, M., Manshadi, F.D., Khalkhali, M., Mousavi, S.J., Baghban, A.A., Montazeri, A. et al., 2015, 'Translation and validation of the Persian version of the STarT Back screening tool in patients with nonspecific low back pain', *Manual Therapy* 20(6), 850–854. https://doi.org/10.1016/j.math.2015.04.006

Aebischer, B., Hill, J.C., Hilfiker, R., Karstens, S., 2015, 'German translation and cross-cultural adaptation of the STarT back screening tool', *PLoS One*, 10(7), 1–14. https://doi.org/10.1371/journal.pone.0132068

Airaksinen, O., Brox, J.I., Cedraschi, C., Hildebrandt, J., Klaber-Moffett, J., Kovacs, F. et al., 2006, 'Chapter 4: European guidelines for the management of chronic non-specific low back pain', *European Spine Journal* 15(Suppl.2), 192–300. https://doi.org/10.1007/s00586-006-1072-1

Alexander, M.C., 2018, *The 11 Languages of South Africa*, viewed 04 January 2019, from http://southafrica-info.com/arts-culture/11-languages-south-africa/.

Beaton, D.E., Bombardier, C., Guillemin, F. & Ferraz, M.B., 2000, 'Guidelines for the process of cross-cultural adaptation of self-report measures', *Spine* 25(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Betten, C., Sandell, C., Hill, J.C. & Gutke, A., 2015, 'Cross-cultural adaptation and validation of the Swedish STarT back screening tool', *European Journal of Physiotherapy* 17(1), 29–36. https://doi.org/10.3109/21679169.2014.1003962

Bier, J.D., Ostelo, R.W.J.G., Hooff, M.L., Koes, B.W. & Verhagen, A.P., 2017, 'Validity and reproducibility of the STarT back tool (Dutch version) in patients with low back pain in primary care settings', *Physical Therapy* 97(5), 561–570. https://doi.org/10.1093/ptj/p2x023

Bolarinwa, O.A., 2015, 'Principles and methods of validity and reliability testing of questionnaires used in social and health science researches', *Nigerian Postgraduate Medical Journal* 22(4), 195–201. https://doi.org/10.4103/1117-1936.173959

Bruyere, O., Demoulin, M., Brereton, C., Humblet, F., Flynn, D., Hill, J.C. et al., 2012, 'Translation validation of a new back pain screening questionnaire (the STarT back screening tool) in French', *Archives of Public Health* 70(12), 1–5. https://doi.org/10.1186/0778-7367-70-12

Glenn, S., 2018, *Test-Retest reliability*, viewed 20 September 2018, from http://www.statisticshowto.com/test-retest-reliability.

Global Burden of Disease Study, 2015, 'Disease and injury incidence and prevalence collaborators, 2016, global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015, 1990–2015', 388(10053), 1545–1602. https://doi.org/10.1016/S0140-6736(16)31678-6

Glover, T.A. & Albers, C.A., 2007, 'Considerations for evaluating universal screening assessments', *Journal of School Psychology* 45(2), 117–135. https://doi.org/10.1016/j.jsp.2006.05.005

Gusi, N., Del Pozo-Cruzo, B., Olivares, P.R., Hernandez-Mocholi, M. & Hill, J.C., 2011, 'The Spanish version of the "STarT back screening tool" (SBST) in different subgroups', *Atencion Primaria* 43(7), 356–361. https://doi.org/10.1016/j.aprim.2010.05.019

Hayden, J.A., Dunn, K.M., Van der Windt, D.A. & Shaw, W.S., 2010, 'What is the prognosis of back pain?', *Best Practice and Research: Clinical Rheumatology* 24(2), 167–179. https://doi.org/10.1016/j.berh.2009.12.005

Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E. et al., 2008, 'A primary care back pain screening tool: Identifying patient subgroups for initial treatment', *Arthritis Care and Research* 59(5), 632–641. https://doi.org/10.1002/art.23563

Hill, J.C., Whitehurst, D.G.T., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E. et al., 2011, 'Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial', *The Lancet* 378(9802), 1560–1571. https://doi.org/10.1016/S0140-6736(11)60937-9

Karstens, S., Krug, K., Hill, J.C., Stock, C., Steinhauser, J., Szecsenyi, J. et al., 2015, 'Validation of the German version of the STarT back tool (STarT-G): A cohort study with patients from primary care practices', *BMC Musculoskeletal Disorders* 16, 1–8. https://doi.org/10.1186/s12891-015-0806-9

Keele University, 2017, *STarT back screening tool*, viewed 15 March 2017, from www.keele.ac.uk/sbst/startbacktool/.

Koes, B.W., Van Tulder, M.W., Lin, C.W., Macedo, L.G., McAuley, J. & Maher, C., 2010, 'An updated overview of clinical guidelines for the management of non-specific low back pain in primary care', *European Spine Journal* 19, 2075–2094. https://doi.org/10.1007/s00586-010-1502-y

Louw, Q.A., Morris, L.D. & Grimmer-Somers, K., 2007, 'The prevalence of low back pain in Africa: A systematic review', *BMC Musculoskeletal Disorders* 8, 1–14. https://doi.org/10.1186/1471-2474-8-105

Luan, S., Min, Y., Li, X., Wu, S. & Hill, J.C., 2014, 'Cross-cultural adaptation, reliability, and validity of the Chinese version of the STarT back screening tool in patients with low back pain', *Spine* 39(16), E974–E979.

Maggiani, A. & Abenavoli, A., 2019, 'Italian translation and cross-cultural adaptation of a back-pain screening questionnaire (STarT back screening tool)', *Annali di Igiene* 31, 69–75. https://doi.org/10.7416/ai.2019.2260

Maher, C., Underwood, M. & Buchbinder, R., 2017, 'Non-specific low back pain', *The Lancet* 389, 736–747. https://doi.org/10.1016/S0140-6736(16)30970-9

Matsudaira, K., Oka, H., Kikuchi, N., Haga, Y. & Sawada, T., 2016, 'Psychometric properties of the Japanese version of the STarT back tool in patients with low back pain', *PLoS One* 11(3), e0152019. https://doi.org/10.1371/journal.pone.0152019

Melloh, M., Elfering, A., Egli Presland, C., Roeder, C., Barz, T., RolliSalath, C. et al., 2009, 'Identification of prognostic factors for chronicity in patients with low back pain: A review of screening instruments', *International Orthopaedics* 33(2), 301–313. https://doi.org/10.1007/s00264-008-0707-8

Morgan, S., 2017, *The Importance of languages*, viewed 18 April 2017, from http://www.importanceoflanguages.com

Morris, L.D., Grimmer-Somers, K.A., Louw, Q.A. & Sullivan, M.J., 2012, Cross-cultural adaptation and validation of the South African Pain Catastrophizing Scale (SA-PCS) among patients with fibromyalgia', *Health and Quality of Life Outcomes* 10(137), 1–12. https://doi.org/10.1186/1477-7525-10-137

Morso, L., Albert, H., Kent, P., Manniche, C. & Hill, J.C., 2011, 'Translation and discriminative validation of the STarT back screening tool into Danish', *European Spine Journal* 20(12), 2166–2173. https://doi.org/10.1007/s00586-011-1911-6

Mphahlele, N., Mitchell, D. & Kamerman, P., 2008, 'Validation of the Wisconsin brief pain questionnaire in a multilingual South African population', *Journal of Pain and Symptom Management* 36(4), 396–412. https://doi.org/10.1016/j.jpainsymman.2007.10.020

Nicholas, M.K., Linton, S.J., Watson, P.J., Main, C.J. & the 'Decade of the Flags' Working Group, 2011, 'Early identification and management of psychological risk factors ('yellow flags') in patients with low back pain: A reappraisal', *Physical Therapy* 91(5), 737–754. https://doi.org/10.2522/ptj.20100224

Nielsen, J., 2012, *Usability 101: Introduction to usability*, viewed 27 November 2018, from http://www.nngroup.com/articles/usability-101-introduction-to-usability/.

Piironen, S., Paananen, M., Haapea, M., Hulpi, M., Zitting, P., Ryynanen, K. et al., 2016, 'Transcultural adaption and psychometric properties of the STarT back screening tool among Finnish low back pain patients', *European Spine Journal* 25, 287–295. https://doi.org/10.1007/s005863-015-3804–3806

Pilz, B., Vasconcelos, R.A., Marcondes, F.B., Lodovichi, S.S., Mello, W. & Grossi, D., 2014, 'The Brazilian version of STarT back screening tool – Translation, cross-cultural adaptation and reliability', *Brazilian Journal of Physical Therapy* 18(5), 453–461. https://doi.org/10.1097/BRS.0000000000000413

Polit, D.F. & Beck, C.T., 2006, 'The content validity index: Are you sure you know what's being reported? Critique and recommendations', *Research in Nursing and Health* 29, 489–497. https://doi.org/10.1002/nur

Raimundo, A., Parraca, J.A., Batalha, N., Tomas-Carus, P., Hill, J.C. & Gusi, N., 2017, 'Portuguese translation cross-cultural adaptation and reliability of the questionnaire "STarT back screening tool" (SBST)', *Acta Reumato Port* 42, 38–46.

Robinson, H.S. & Dagfinrud, H., 2017, 'Reliability and screening ability of the STarT back screening tool in patients with low back pain in physiotherapy practice, a cohort study', *BMC Musculoskeletal Disorders* 18, 1–7. https://doi.org/10.1186/s12891-017-1553-x

Sharafi, S.E., Hafizi, S., Shahi, M.H., Kordi, R., Noorbala, A.A., Arbabi, M. et al., 2017, 'The Persian version of the Orebro musculoskeletal pain screening questionnaire: Translation and evaluation of its psychometric properties', *International Journal of Preventative Medicine* 8(14), 1–12. https://doi.org/10.4103/2008-7802.201658

Terwee, C.B., Bot, S.D.M., De Boer, M.R., Van der Windt, D.A.W.M., Knol, D.L., Dekker, J. et al., 2007, 'Quality criteria were proposed for measurement properties of health status questionnaires', *Journal of Clinical Epidemiology* 60(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Tsang Chi Chung, R., Lau Sau Ying, J., Kwang So Fong, S., So Ming Loi, E., Law Ka Yee, R., Wong Fu Yan, T. et al., 2017, 'Reliability, construct and predictive validity of the Hong Kong Chinese Orebro musculoskeletal pain screening questionnaire', *Journal of Occupational Rehabilitation* 24(4), 584–592. https://doi.org/10.1007/510926-016-9689-3

United Kingdom (UK) Essays, 2016, *The effects of apartheid on South African education politics essay*, viewed 06 March 2018, from http://www.ukessays.com/essays/politics/the-effects-of-apartheid-on-the-south-african-education-politics-essay.php.

United Nations Development Program (UNDP), 2017, *Income inequality trends in sub-Saharan Africa*, viewed 06 March 2018, from http://www.africa.undp.org/content/dam/rba/docs/Reports/undp-rba_Income%20Inequality%20in%20SSA_Infographic%20Booklet_EN.pdf.

United Nations Population Division, 2017, *World population prospects*, viewed 02 February 2019, from viewed from https://population.un.org/wpp/DataQuery/.

Van Dyke, J.A., 2016, 'Understanding the mechanisms of language comprehension, viewed 22 November 2018, from http://www.researchfeatures.com/2016/11/22/understanding-mechanisms-language-comprehension/html.

Van Tulder, M., Becker, A., Bekkering, T., Breen, A., Del Real, M.T., Hutchinson, A. et al., 2006, 'Chapter 3, European guidelines for the management of acute nonspecific low back pain in primary care', *European Spine Journal* 15(S2), S169–S191. https://doi.org/10.1007/s0586-006-1071-2

Yaghmaie, F., 2003, 'Content validity and its estimation', *Journal of Medical Education* 3(1), 25–27. https://doi.org/10.22037/JME.V311.870

Yoshimoto, T., Oka, H., Katsuhira, J., Fujii, T., Masuda, K., Tanaka, S. et al., 2017, 'Prognostic psychosocial factors for disabling low back pain in Japanese hospital workers', *PLoS One* 12(5), 1–12. https://doi.org/10.1371/journal.pone.0177908