Validation and reliability of a Japanese version of the Simple Shoulder Test: a cross-sectional study

Takuya Sekiguchi, MD, PhD [a,b], Yoshihiro Hagiwara, MD, PhD [b,*], Akira Ando, MD, PhD [c], Kenji Kanazawa, MD, PhD [d], Kazuaki Suzuki, MD [b], Masashi Koide, MD, PhD [c], Yutaka Yabe, MD, PhD [b], Satsuki Onoda, MD, PhD [e], Eiji Itoi, MD, PhD [b]

[a] Department of Orthopedic Surgery, JR Sendai Hospital, Sendai, Japan
[b] Department of Orthopedic Surgery, Tohoku University School of Medicine, Sendai, Japan
[c] Department of Orthopedic Surgery, Matsuda Hospital, Sendai, Japan
[d] Department of Orthopedic Surgery, Miyagi South Medical Center, Sendai, Japan
[e] Department of Orthopedic Surgery, Iwate Prefectural Central Hospital, Morioka, Japan

Background: The Simple Shoulder Test (SST) is a widely used patient-reported outcome assessment. The purpose of this study was to develop and validate a Japanese version of the SST (SST-Jp).
Methods: A two-stage observational study was conducted to validate the cross-cultural adaptation of the SST. A total of 100 patients with shoulder disorders completed the SST-Jp; the Disability of Arm, Shoulder, and Hand assessment; and the Medical Outcomes Short-Form 36 (SF-36) at an initial visit. Thirty-four of the patients repeated the SST-Jp one week after the first examination. The test-retest reliability was quantified using the interclass correlation coefficient, and Cronbach's alpha (α) was calculated to assess internal consistency. Construct validity was assessed using Spearman's rank correlation coefficient.
Results: The internal consistency of the SST-Jp was very high ($\alpha = 0.826$). The interclass correlation coefficient of the SST-Jp was also high (0.859). There was a strong, positive correlation between the Disability of Arm, Shoulder, and Hand and the SST-Jp ($r = 0.717$, $P < .001$). The SST-Jp was significantly correlated with most of the SF-36 subscales. The correlations of the SST-Jp with physical subscales of the SF-36 were stronger than those with the other subscales.
Conclusions: The SST-Jp was found to be a valid and reliable measurement for shoulder joint pain and function assessment among the Japanese population.

Shoulder disorders, such as rotator cuff tear and frozen shoulder, are the second most common musculoskeletal disorders.[3] The prevalence of shoulder pain ranges from 6.9% to 26% in the general population,[17] and more specifically, the prevalence of neck and shoulder pain ranges from 10% to 48% in Japan.[10,25,26] Although shoulder disorders are not life-threatening, they can limit the activities of daily living (ADLs) and negatively impact quality of life.[15] Furthermore, shoulder disability may disrupt psychological and social well-being.[22]

Patient-reported outcomes (PROs) have been widely used in clinical practice and research to evaluate treatments to reduce the discrepancy between clinical evaluation and patient perception. There are many different PROs to evaluate shoulder conditions, such as the Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH);[12] Western Ontario Shoulder Instability Index;[14] Shoulder Pain and Disability Index;[23] and Simple Shoulder Test (SST).[16] The SST is one of the most commonly used instruments, in part, because it is a simple and short PRO tool. It has been tested in various clinical settings[6,24] and can be completed within 3 minutes, and patients can complete it before the clinical visit, making it very practical.[16] For proper use of the SST in non−English-speaking populations, it is necessary to use a translated version of the SST that is culturally adapted and properly evaluated. The need for validated translations has become more essential with the growing number of multicenter and multinational studies, which provide more statistical power for randomized controlled trials.[2,9] The SST

has been translated into several languages.[5,18,20,21,28] Although a Japanese version of the SST (SST-Jp) has not been developed and validated, studies using the SST in Japanese populations without cultural adaptation and validation have been published. Cultural adaptation and validation of the SST would benefit Japanese clinicians, surgeons, and patients by improving the rigor of and evidence for a standardized SST-Jp.

Therefore, the purpose of this study was to translate and culturally adapt the SST for use in Japan and evaluate its validity and reliability in Japanese patients with shoulder pain and dysfunction.

Materials and methods

This study was conducted in two phases. In phase 1, cross-cultural adaptation was conducted according to guidelines of the American Association of Orthopaedic Surgeons.[2] In phase 2, we assessed the measurement properties in a Japanese population.

Cross-cultural adaptation

The SST-Jp was developed according to the American Association of Orthopaedic Surgeons guidelines for cross-cultural adaptation.[2] The process consisted of five steps, each of which was documented with a written report. First, the English version of the SST was translated into Japanese by two bilingual translators (forward translation). One of the translators was an experienced shoulder surgeon and aware of the concept of the translation, while the other was a professional translator without a medical background who was not aware or informed of the medical concepts. For step 2, the two translations were synthesized into one after discussion with the study team. In step 3, the synthesized Japanese version was translated into English by two professional translators (back translation). In step 4, any inconsistencies between the original and the back-translated English version were resolved in the second draft of the questionnaire. Finally, all versions were evaluated and a pilot test was conducted. For the pilot study, the SST-Jp was administered to 30 patients with shoulder dysfunction. After analysis of the pilot test, results were found to be acceptable, and the final version of the SST-Jp was confirmed (Supplemental Appendix).

Patients

We recruited patients with shoulder disorders who were admitted to one of four hospitals between February and October 2019. Patients who were 18 years or older, native Japanese speakers, and able to complete the questionnaire by themselves were included in the study. The exclusion criteria were a previous major trauma (e.g., fracture and dislocation) on the affected shoulder or girdle; impairment in the cervical spine, elbow, and/or hand affecting the shoulder function; a history of inflammatory arthritis, physical and/or mental dysfunction as an underlying disease; or a regional tumor and metastasis. Patients completed the following questionnaires: the SST-Jp, the official Japanese version of the 36-item short-form health survey (SF-36 v2),[7,8] and the Japanese Society for Surgery of the Hand version of the DASH.[13] Basic characteristics such as sex, age, affected side, and diagnosis were also recorded. A series of 100 patients with shoulder disorders were included in the study. Among the patients, 36 were re-administered the SST-Jp after one week. These were surgical patients who returned for a preoperative appointment and did not have additional treatments in the previous week. This study was approved by the institutional review board of the authors' affiliated institutions. All patients provided written, informed consent before participating in this study.

Assessments

The SST is a shoulder function scale consisting of 12 questions with 'yes' or 'no' answers to assess the ability to tolerate or perform ADLs. The SST scores range from 0 (worst) to 100 (best) reported as the percentage of answered items to which the patient responds in the affirmative.[1] The SF-36 is a widely used instrument that measures 8 subscales of health including physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality, social functioning (SF), role emotional, and mental health. Each subscale score was transformed into a 0 to 100 scale where a lower score indicated worse quality of life and more severe disability.[7,8] The DASH score is a general upper extremity score used for patients with any condition related to the joints of the upper extremities. It consists of 30 items with a total score ranging from 0 (no disability) to 100 (worst).[13]

Statistical analyses

Continuous variables are summarized as means ±standard deviation. Categorical variables are summarized as numbers (%). Floor and ceiling effects were evaluated when 15% or more of the subjects achieved the lowest (0) or highest (100) possible points on the SST-Jp.[19] Internal consistency was assessed by Cronbach's alpha (α), which was calculated at an anticipated value ranging between 0.65 and 0.95.[4] Intraclass correlation coefficients (ICCs) were calculated to assess the test-retest reliability where values > 0.7 were assumed to sufficiently support test-retest reliability.[27] Measurement error is expressed as the standard error of measurement (SEM) or the minimal detectable change (MDC). The SEM was calculated using the square root of the variance between the measurements and the error variance of the ICC. The MDC was calculated as the SEM*$1.96\sqrt{2}$.

The DASH and SF-36 scores were used to assess the construct validity of the SST-Jp. Construct validity was evaluated using Spearman's rank correlation coefficients because they do not rely on normal distributions (using the Kolmogorov-Smirnov test). Correlation coefficients were categorized as very weak or negligible (below 0.20), weak or low (between 0.20 and 0.40), moderate (between 0.40 and 0.70), strong or high (between 0.70 and 0.90), and very strong or very high (above 0.90).[11] All tests were 2-tailed, and $P < .05$ was considered statistically significant. All statistical analyses were performed using SPSS, version 24.0 (SPSS Japan Inc., Tokyo, Japan).

Results

Demographic data

The mean age of the participants was 64.3 years (range: 56 to 72 years), and 46% (n = 46) were women. In 53 cases, pain and/or dysfunction was in the right shoulder. The disorders were classified as rotator cuff tear in 59 cases, frozen shoulder in 25 cases, shoulder periarthritis in 9 cases, osteoarthritis in 2 cases, and "other" in 5 cases (Table I). The SST-Jp showed no missing data. In total, 3% of the participants reported the worst possible score (0 points) and 5.0% reported the best possible score (100 points).

Reliability

Table II summarizes the reliability of the SST-Jp. The internal consistency for the 12 questions in the SST-Jp was high ($\alpha = 0.826$). The ICC of the SST-Jp was 0.859 (95% confidence interval: 0.715–0.930; MDC = 3.39%).

Table I
Baseline characteristics of the participants

Characteristics (n = 100)	Total
Sex, %	
Male	54.0
Female	46.0
Age (years), mean (SD)	64.3 (10.5)
Affected side, %	
Right	53.0
Left	47.0
Diagnosis, %	
Rotator cuff tear	59.0
Frozen shoulder	25.0
Periarthritis	9.0
Osteoarthritis	2.0
Others	5.0

SD, standard deviation.

Table III
Correlation of the SST-Jp with the DASH and SF-36 (validity)

Instrument	r (P value)* SST-Jp
DASH	0.717 (P < .001)
SF-36 subscales	
Physical function	0.493 (P < .001)
Role physical	0.545 (P < .001)
Bodily pain	0.545 (P < .001)
General health	0.200 (P = .046)
Vitality	0.223 (P = .025)
Social function	0.345 (P = .005)
Role emotion	0.350 (P < .001)
Mental health	0.307 (P < .001)

SST-Jp, Japanese version of the Simple Shoulder Test, DASH, Disability of Arm, Shoulder and Hand, SF-36, Short-Form 36.
* Calculated using Spearman's rank correlation coefficient of the SST-Jp with DASH and SF-36.

Validity

The validity results of the SST-Jp are shown in Table III. There was a strong correlation between the DASH and the SST-Jp (r = 0.717). The SST-Jp was shown to have a significant correlation with all SF-36 subscales, as follows: a moderate correlation with the PF (r = 0.493), RP (r = 0.545), and BP (r = 0.545) subscales and a weak correlation with the GH (r = 0.200), vitality (r = 0.223), SF (r = 0.279), SF (r = 0.345), role emotional (r = 0.350), and GH (r = 0.307) subscales.

Discussion

This study aimed to culturally adapt and validate the SST for use in Japan. The results show that the SST-Jp presented in this study has high internal consistency and reliability for the subscores and the total score and, therefore, would be a reliable and valid measure of shoulder pain and function that can be implemented in Japanese clinical practice. In addition, the SST-Jp showed a strong correlation with the DASH and moderate correlation with the SF-36 physical subscales, including the PF, RP, and BP.

The values for internal consistency in the present study were similar to those reported from other studies to adapt the SST for different cultural contexts.[5,18,20,21,28] The ICC indicated good reproducibility in the present study, and the value was similar or slightly higher than the values obtained in the other versions of the SST.[5,18,20,21,28] In addition, the construct validity of the SST-Jp was good, as evidenced by a strong correlation between the SST-Jp and DASH, which was similar or higher than that observed in the Dutch, Spanish, and Persian versions of the SST (r = 0.60–0.74).[5,20,28] Although the SST-Jp showed moderate correlation with the physical subscales, such as PF, RP, and BP, they were lower than the correlations observed between the SST-Jp and the DASH. Furthermore, other subscales of the SF-36 showed a weak correlation with the SST-Jp. The SST-Jp was assumed to reflect the pain and function of the shoulder better than the general health. These results are consistent with the results obtained with the Persian version of the SST.[5]

One strength of this study was the standardized method for cross-cultural adaptation. Furthermore, the participants were recruited from 4 medical centers, which reduced potential selection bias. However, this study has several limitations to consider. First, owing to the cross-sectional design, the sensitivity to change (responsiveness) and the minimal clinical difference in responsiveness (perceptibility) were not analyzed. Second, the mean age of this study's participants was slightly high. Third, a sample size calculation was not performed before study recruitment to determine the optimal sample size for robust statistical analysis. Although 100 patients can be considered an excellent sample size, 36 patients for the test-retest reliability assessment is only moderately good.[27] It would be necessary to increase the number of patients, especially to strengthen the test-retest reliability outcome.

Conclusion

The proposed SST-Jp has satisfactory internal consistency, test-retest reliability, and construct validity. Therefore, we recommend the SST-Jp for evaluating patients with shoulder dysfunction in the Japanese population.

Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jseint.2020.10.018.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Table II
Cronbach's alpha and ICC of the SST-Jp (reliability)

Instrument	Cronbach's alpha	ICC	95% CI
SST-Jp	0.826	0.859	0.715–0.930

SST-Jp, Japanese version of the Simple Shoulder Test, ICC, intraclass correlation coefficient, CI, confidence interval.

References

1. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the arm, shoulder, and hand questionnaire (DASH) and its short version (QuickDASH), shoulder pain and disability index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, constant (Murley) score (CS), simple shoulder test (SST), Oxford shoulder score (OSS), shoulder disability

questionnaire (SDQ), and Western Ontario shoulder instability index (WOSI). Arthritis Care Res (Hoboken) 2011;63:S174. https://doi.org/10.1002/acr.20630.

2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186-91.

3. Burbank KM, Stevenson JH, Czarnecki GR, Dorfman J. Chronic shouler pain part I: Evaluation and diagnosis. Am Fam Physician 2008;77:453-60.

4. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.

5. Ebrahimzadeh MH, Vahedi E, Baradaran A, Birjandinejad A, Seyyed-Hoseinian SH, Bagheri F, et al. Psychometric properties of the Persian version of the Simple Shoulder Test (SST) Questionnaire. Arch Bone Jt Surg 2016;4:387-92.

6. Ekeberg OM, Bautz-Holter E, Tveita EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. BMC Musculoskelet Disord 2008;9:68. https://doi.org/10.1186/1471-2474-9-68.

7. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol 1998;51:1037-44.

8. Fukuhara S, Ware JE Jr, Kosinski M, Wada S, Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. J Clin Epidemiol 1998;51:1045-53.

9. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.

10. Hagiwara Y, Sekiguchi T, Yabe Y, Sugawara Y, Watanabe T, Kanazawa K, et al. Living status, economic hardship and sleep disturbance were associated with subjective shoulder pain in survivors of the Great East Japan Earthquake: A cross sectional study. J Orthop Sci 2017;22:442-6. https://doi.org/10.1016/j.jos.2016.12.027.

11. Haldorsen B, Svege I, Roe Y, Bergland A. Reliability and validity of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire in patients with shoulder impingement syndrome. BMC Musculoskelet Disord 2014;15:78. https://doi.org/10.1186/1471-2474-15-78.

12. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996;29:602-8.

13. Imaeda T, Toh S, Nakao Y, Nishida J, Hirata H, Ijichi M, et al. Validation of the Japanese Society for Surgery of the Hand version of the Disability of the Arm, Shoulder, and Hand questionnaire. J Orthop Sci 2005;10:353-9. https://doi.org/10.1007/s00776-005-0917-5.

14. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. Clin J Sport Med 2003;13:84-92. https://doi.org/10.1097/00042752-200303000-00004.

15. Largacha M, Parsons IMt, Campbell B, Titelman RM, Smith KL, Matsen F 3rd. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: a study of 2674 patients. J Shoulder Elbow Surg 2006;15:30-9. https://doi.org/10.1016/j.jse.2005.04.006.

16. Lippitt SB, Harryman DT, Matsen FA, Lippitt S, Lippit S. A practical tool for evaluating function: The simple shoulder test. The shoulder: a balance of mobility and stability. Rosemont (IL): American Academy of Orthopaedic Surgeons; 1993.

17. Luime J, Koes B, Hendriksen I, Burdorf A, Verhagen A, Miedema H, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. Scand J Rheumatol 2004;33:73-81. https://doi.org/10.1080/03009740310004667.

18. Marchese C, Cristalli G, Pichi B, Manciocco V, Mercante G, Pellini R, et al. Italian cross-cultural adaptation and validation of three different scales for the evaluation of shoulder pain and dysfunction after neck dissection: University of California-Los Angeles (UCLA) Shoulder Scale, Shoulder Pain and Disability Index (SPADI) and Simple Shoulder Test (SST). Acta Otorhinolaryngol Ital 2012;32:12.

19. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293-307.

20. Membrilla-Mesa M, Tejero-Fernández V, Cuesta-Vargas A, Arroyo-Morales M. Validation and reliability of a Spanish version of Simple Shoulder Test (SST-Sp). Qual Life Res 2015;24:411-6. https://doi.org/10.1007/s11136-014-0760-2.

21. Neto JOB, Gesser RL, Steglich V, Ferreira APB, Gandhi M, Vissoci JR, et al. Validation of the Simple Shoulder Test in a Portuguese-Brazilian population. Is the latent variable structure and validation of the Simple Shoulder Test Stable across cultures? PLoS One 2013;8:e62890. https://doi.org/10.1371/journal.pone.0062890.

22. Paananen M, Taimela S, Auvinen J, Tammelin T, Zitting P, Karppinen J. Impact of self-reported musculoskeletal pain on health-related quality of life among young adults. Pain Med 2011;12:9-17. https://doi.org/10.1111/j.1526-4637.2010.01029.x.

23. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Rheum 1991;4:143-9.

24. Schmidt S, Ferrer M, González M, González N, Valderas JM, Alonso J, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg 2014;23:434-44. https://doi.org/10.1016/j.jse.2013.09.029.

25. Sekiguchi T, Hagiwara Y, Sugawara Y, Tomata Y, Tanji F, Watanabe T, et al. Influence of subjective economic hardship on new onset of neck pain (so-called: katakori) in the chronic phase of the Great East Japan Earthquake: A prospective cohort study. J Orthop Sci 2018;23:758-64. https://doi.org/10.1016/j.jos.2018.04.011.

26. Takasawa E, Yamamoto A, Kobayashi T, Tajika T, Shitara H, Ichinose T, et al. Characteristics of neck and shoulder pain in the Japanese general population. J Orthop Sci 2015;20:403-9. https://doi.org/10.1007/s00776-014-0676-2.

27. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651-7. https://doi.org/10.1007/s11136-011-9960-1.

28. van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg 2012;21:808-14. https://doi.org/10.1016/j.jse.2011.09.026.