Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

CLINICAL RESEARCH

Crosscultural Adaptation and Validation of the Korean Version of the New Knee Society Knee Scoring System

Seok Jin Kim MD, Mohnish Singh Basur DNB (ortho), Chang Kyu Park MD,
Suri Chong MD, Yeon Gwi Kang MS, Moon Ju Kim MS,
Jeong Seong Jeong MS, Tae Kyun Kim MD, PhD

Abstract

Background The 2011 Knee Society Score© (2011 KS Score©) is used to characterize the expectations, symptoms, physical activity, and satisfaction of patients who undergo TKA and is widely used to assess the outcome of TKA. However, it has not been adapted or validated for use in Korea.

Questions/purposes We developed a Korean version of the 2011 KS Score and evaluated the (1) test-retest reliability, (2) convergent validity, and (3) responsiveness of the Korean version.

Methods The Korean version of the 2011 KS Score was derived by using a well-established translational procedure based on international guidelines, which include translation, synthesis, back-translation, expert committee review, pretesting, and submission for appraisal. A total of 123 patients with knee osteoarthritis who were scheduled to undergo TKA were recruited for the study. Ninety percent of the patients (111 of 123) were women, which is an exact representation of the Korean population having TKAs. To evaluate reliability, the patients were evaluated twice during a 4-week interval using the questionnaire. Reliability was assessed by using intraclass correlation coefficients (ICCs) and internal consistency by using Cronbach's alpha to determine the validity of the Korean version of the 2011 KS Score. The patients were evaluated by using the validated Korean versions of the WOMAC and SF-36 questionnaires. Spearman's correlation coefficient was used for validation. Responsiveness was determined by calculating the standardized response mean from the preoperative and postoperative test scores in

S. J. Kim
Department of Orthopaedic Surgery, Chamjoeun Hospital, Gwangju, Korea

M. S. Basur
Department of Orthopaedic Surgery, Al Raha Hospital, AbeerGroup, Abu Dhabi, UAE

C. K. Park, S. Chong, Y. G. Kang, M. J. Kim, J. S. Jeong, T. K. Kim (✉)
Joint Reconstruction Center, Seoul National University Bundang Hospital, 82 Gumi-ro, 173 Beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea
e-mail: osktk@snubh.org

T. K. Kim
Department of Orthopaedic Surgery, Seoul National University College of Medicine, Seoul, Republic of Korea
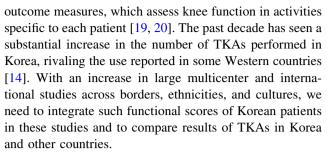
🖄 Springer

the Korean version of the 2011 KS Score. To address the gender disparity in our study we identified 53 males who underwent TKA for osteoarthritis after completion of this study and generated age-matched controlled groups to evaluate construct validity and responsiveness in Korean males.

Results   The reliability proved good to excellent with an ICC between 0.69 and 0.85, depending on the clinical properties tested, which included the following: symptoms, satisfaction, expectation, and total functional activity consisting of functional activity, standard activity, advanced activity, and discretionary activity. All subscales showed good to excellent internal consistency indicated by Chronbach's alpha (range, 0.83–0.92). For validity, three of the four domains (the exception was expectation) of the 2011 KS Score, correlated either strongly or moderately with the Korean WOMAC score (r ≥ 0.35). When compared with the SF-36, the satisfaction domain showed a weak positive correlation with all the subscales of the SF-36 except general health (r < 0.35). The activity domain showed a strong positive correlation with physical function (r = 0.62) and physical component summary (r = 0.52), moderate with physical role (r = 0.46), and weak with bodily pain (r = 0.26) and social function (r = 0.31). The symptom domain also exhibited a similar moderate positive correlation with physical function (r = 0.41) and weak positive correlation with bodily pain, social function, and physical component summary (r = 0.22, 0.20, and 0.26, respectively). For responsiveness, all the domains of Korean version of the 2011 KS Score, except for expectation, showed large changes (> 0.8), calculated as standardized response mean. The total amount of the Korean version of the 2011 KS Score (2.03, p < 0.001) showed higher responsiveness when compared with the WOMAC total (1.88, p < 0.001) and SF-36 physical and mental component summaries (1.14, p < 0.001; and 0.68, p < 0.001, respectively).

Conclusions   The Korean version of the 2011 KS Score was successfully developed using a process of crosscultural adaptation for the Korean-speaking population who had undergone TKA for osteoarthritis of the knee. The Korean version of the 2011 KS Score was shown to be a reliable, valid, and responsive tool and can be used to assess functional outcomes and expectations of Korean patients who undergo TKA. The demographic features of TKA in the Korean population should be taken into account with additional studies recommended to further investigate these psychometric properties in Korean men.

Level of Evidence   Level II, diagnostic study.

Introduction

Evaluation of a patient's functional outcome after TKA is challenging and requires an objective and a functional assessment. Greater importance has been attached to functional scoring systems in the form of patient-reported

outcome measures, which assess knee function in activities specific to each patient [19, 20]. The past decade has seen a substantial increase in the number of TKAs performed in Korea, rivaling the use reported in some Western countries [14]. With an increase in large multicenter and international studies across borders, ethnicities, and cultures, we need to integrate such functional scores of Korean patients in these studies and to compare results of TKAs in Korea and other countries.

The Knee Society Clinical Rating System was developed in 1989 as an objective scoring system to assess patient function and rapidly became a popular method for reporting outcomes after TKA [13, 20]. In 2011, the new Knee Society Knee Scoring System© (2011 KS Score©) [19] was developed, which consists of objective items from the previous Knee Society scores and completely new subjective items in the form of patient-reported outcome measures. The 2011 KS Score has been validated, confirming its overall reliability and consistency and those of its different domains, and has been reported to be without the ambiguities of the previous scoring systems [19]. However, to our knowledge, the 2011 KS Score has not been adapted and validated for use in Korean patients undergoing knee surgery. This scoring system not only should be translated, but also culturally adapted to the native Korean-speaking population through the process of crosscultural adaptation.

The purpose of this study therefore was to establish a Korean version of the 2011 KS Score for Korean-speaking patients who undergo TKA, developed through crosscultural adaptation and to investigate the psychometric properties of the Korean version of the 2011 KS Score in terms of: (1) test-retest reliability, which refers to the degree to which test results are consistent over a brief internal consistency, defined as coherence among the different scale components, without any treatment change; (2) construct validity, defined as the degree to which a test measures what it claims or purports to be measuring when compared with scores with proven validity such as the Korean WOMAC and Korean SF-36; and (3) responsiveness, defined as the ability to reflect changes in patient status by comparing preoperative and postoperative results.

Materials and Methods

Translation Procedure

Korean translation was performed using guidelines provided by Guillemin et al. [8, 9] using the crosscultural adaptation process. This process not only ensures an appropriate linguistically translated version, but also adapts to maintain content validity across cultures. The

crosscultural adaptation process is conducted in six stages: translation, synthesis, back-translation, expert committee review, pretesting, and submission for appraisal. Briefly, the English version of the Knee Society Score was translated separately by three native Korean bilingual translators. After uniform agreement was reached among the three translators, a pretest Korean translation version was established. This version was back-translated by two bilingual native English speakers who were blinded to the original English version. We continued this process until a final version was produced that had no disagreements between the English and Korean versions. When the consensus version was formed, the back-translated English version was sent to and approved by the inventor of the Knee Society Score. The final version was pretested in 20 Korean patients with knee osteoarthritis.

Study Design

A total of 350 patients were requested to participate in our study, from which 123 patients met the inclusion exclusion criteria and were available for final analysis (Fig. 1). All

patients had knee osteoarthritis and were scheduled to undergo TKA at our institute between June 2013 and February 2014. Of the patients, 90% (111 of 123) were females; to address this issue of gender disparity we later added 53 males to this mixed gender population as a separate analysis. The mean BMI was 27 kg/m^2 (mean ± SD, 4 kg/m^2) (Table 1). All patients with knee osteoarthritis undergoing TKA and available to undergo 1-year postoperative followup were included in the study. Patients with hip or spine disease, congenital deformity, knee infection, history of surgery on the ipsilateral or contralateral leg, and those who declined to participate in the study were excluded. Patients who had had a second, contralateral TKA within this 12-month period also were excluded to ensure that the results reflected the outcome of the index operation and not a subsequent operation. The participants were informed properly about the study and signed the consent form.

To address the issue of gender disparity in our study, after completion of our study we identified 71 males with knee osteoarthritis who underwent TKA at our institute between February 2014 and December 2016 from our patient database (as per our routine followup protocol all

Fig. 1 The flowchart shows patient enrollment and exclusion criteria in our study. * = patients not meeting inclusion criteria.

Total TKAs During the Study Period

350 Patients

*Patients Not Meeting Inclusion Criteria - 166

Patients Lost to Followup - 29

Patients Refusing to Participate in Study - 32

*Concurrent Spine or Hip Problem - 57

History of Surgery on Ipsilateral or Contralateral Leg - 23

Patients With Congenital Deformity of Knee - 2

Patients With Prior Knee Sepsis - 1

Patients Who Had a Second, Contralateral TKA Within This 12-month Period - 83

Final Analysis

123 patients

Table 1. Demographic data of the total patient population

Variable	Total population (n = 123)
Age (years, mean ± SD)	71.0 ± 6.0
Sex, number (%)	
Women	111 (90%)
Men	12 (10%)
Side, number (%)	
Right	53 (43%)
Left	70 (57%)
Height (cm, mean ± SD)	153 ± 7
Weight (kg, mean ± SD)	64 ± 10
BMI (kg/m^2, mean ± SD)	27 ± 3

patients complete three questionnaires). Of these, 53 met the same inclusion and exclusion criteria as study participants, with at least 1 year followup. We generated age-matched control groups in the ratio of 2:1 with these 53 men and 106 women from the earlier mixed-gender population to comparatively evaluate validity and responsiveness of the Korean New KSS in Korean men. Although we could not provide comparative reliability data, data of validity and responsiveness are enough to resolve the concern surrounding gender composition of our study. Approval for this study was obtained from the ethical and review boards of our institute.

Questionnaires

All patients were asked to complete a questionnaire containing the Korean-translated New Knee Society Score (KSS), Korean WOMAC, and Korean SF-36. All the questionnaires initially were filled out by patients and were checked for missing and appropriate responses by a research assistant (JSJ) during subsequent visits, in the presence of the patient. A research assistant was employed to limit responder burden, because patients were being asked to simultaneously complete different questionnaires containing many similar questions. Clinical examination and radiologic assessment were performed by a trained fellow (SJK) in knee arthroplasty following a written protocol so that all the patients were examined using the same method. All personnel involved in data collection were trained by the lead author (TKK) and used a standardized method of data collection following a written protocol to minimize interobserver variability as much as possible. Written guidelines regarding how to assess pain also were developed and given to all the investigators. Data collection was started 4 weeks (Korean New KSS, Korean WOMAC, and Korean SF-36) and 1 day before the

operation (Korean New KSS) and 1 year after the operation (Korean New KSS, Korean WOMAC, and Korean SF-36).

To evaluate the test-retest reliability of the Korean-translated New KSS, all patients completed the questionnaires 4 weeks apart without any intervening treatment (4 weeks, 1 day before the operation) with the assumption that during this period, the spontaneous change in the condition of their osteoarthritis was minimal. Reliability was assessed by using intraclass correlation coefficients (ICCs) with 95% CIs [4, 10]. Internal consistency, defined as coherence among the different scale components, was assessed by using Cronbach's alpha coefficient.

The ideal method of measuring validity is to compare it with a gold standard test; however, currently, no gold standard test has been established that perfectly reflects pre- and post-TKA status [5, 7]. Owing to the lack of such a test, the domains of the Korean-translated New KSS were tested by comparing them with the appropriate subscales of the Korean WOMAC and Korean SF-36 by using Spearman's coefficient. These scoring systems were selected because their test validities have been proven in previous studies [2, 11].

Responsiveness is a nonstatistical term that shows the ability to reflect changes in preoperative and postoperative results, in which the higher the responsiveness, the greater the ability to detect changes. Responsiveness was evaluated by comparing preoperative and postoperative scores among the Korean New KSS, Korean WOMAC, and Korean SF-36 by using the standardized response means (SRM). These scoring systems were selected based on prior rigorous psychometric evaluations [2, 11].

Statistical Analyses

An ANOVA was used to calculate ICC with 95% CI between the first and second applications of the Korean New KSS to assess test-retest reliability. An ICC greater than 0.8 correlates with excellent reproducibility. Internal consistency was assessed by using Chronbach's alpha. An alpha of 0.7 was considered fair; 0.8, good; and 0.9, excellent. Construct validity was estimated by using the correlation between the domains of the 2011 KS Score and the other questionnaires (Korean version of the WOMAC and SF-36) with Spearman's coefficient (r). These correlations can be converging (positive) or diverging (negative). Correlation was considered strong if the value was greater than 0.5; moderate if the value was between 0.5 and 0.35; and weak if the value was less than 0.35. Responsiveness was evaluated by comparing the responsiveness of the Korean New KSS with the treatment (TKA) by using the SRM, calculated as the mean difference between the preoperative and 12-month postoperative

scores divided by the SD of the score; the larger the SRM value, the greater is the responsiveness. A SRM of 0.2 to 0.8 reflects a small change; 0.5 to 0.8, a moderate change; and greater than 0.8, a large change. The means and SDs of the SRMs of the two measurements were estimated with a jackknife procedure, and then tested with a paired t-test [15, 18]. Because TKA is effective for reducing patients' pain and improving their quality of life, we assumed a large standardized response mean (> 0.8) for all the scores. The preoperative and postoperative Korean New KSS scores were assessed against the Korean WOMAC and Korean SF-36 questionnaires by using the SRM. All statistical analyses were performed by a trained statistician (YGK) using IBM SPSS, Version 22.0 (IBM Corp, Armonk, NY, USA). All the scores were reported as mean and SD, and p values less than 0.05 were considered statistically significant.

Results

The test-retest reliability and internal consistency proved excellent or good to excellent for all domains of the Korean New KSS. All the domains of the Korean New KSS exhibited an ICC between 0.69 and 0.85, depending on the domains tested, which proves adequate reproducibility. Internal consistency, as indicated by Chronbach's alpha, ranged from 0.83 to 0.92 for the individual subscales and was good to excellent for all the domains (Table 2). All the subscales of the Korean New KSS (symptoms, satisfaction, expectation, and activity) had an ICC greater than 0.8. However, in the activity subscale, only functional activities and discretionary activities showed ICCs of 0.73 and 0.71 and Chronbach's alpha of 0.84 and 0.83, respectively.

The Korean New KSS overall scores correlated well with the Korean WOMAC and Korean SF-36 scores. When compared with the WOMAC, all the domains of the Korean New KSS correlated either strongly or moderately with the individual subscales of Korean WOMAC score except for expectation ($r \geq 0.35$, $p < 0.001$). Furthermore, all the domains of the Korean New KSS correlated with the Korean WOMAC total score (symptom: $r = -0.53$, $p < 0.001$; satisfaction: $r = -0.49$, $p < 0.001$; expectation: $r = 0.21$, $p = 0.019$; activity: $r = -0.53$, $p < 0.001$) (Table 3). When compared with the Korean SF-36, the Korean New KSS showed good correlation of its satisfaction and

Table 2. Test-retest reliability and internal consistency of the total scores

Domain	Test 1 Mean (SD)	Test 2 Mean (SD)	ICC*	95% CI	Cronbach's alpha
Symptom (3 items)/25 points	5.41 (4.38)	5.92 (4.00)	0.85	0.78–0.89	0.92
Satisfaction (5 items)/40 points	12.10 (6.90)	12.88 (7.52)	0.81	0.73–0.86	0.89
Expectation (3 items)/15 points	12.89 (2.19)	13.03 (2.20)	0.80	0.73–0.86	0.89
Total functional activity (19 items)/100 points	30.80 (16.91)	29.92 (17.15)	0.81	0.74–0.87	0.90
Functional activity (5 items)/30 points	13.35 (8.69)	11.59 (9.13)	0.71	0.61–0.79	0.84
Standard activity (6 items)/30 points	9.54 (5.46)	9.54 (5.81)	0.84	0.78–0.89	0.91
Advanced activity (5 items)/25 points	3.29 (3.75)	3.33 (3.93)	0.83	0.77–0.88	0.91
Discretionary activity (3 items)/15 points	4.62 (4.22)	5.46 (4.12)	0.69	0.58–0.78	0.83
Total score/180 points	61.20 (24.18)	62.45 (24.22)	0.86	0.81–0.90	0.93

*Mixed effects ANOVA model with absolute agreement (single measure); ICC = intraclass correlation coefficient.

Table 3. Construct validity between the Korean New KSS and the Korean WOMAC

Korean WOMAC		Korean New KSS (preoperative form)			
Subscales	Score Mean ± SD	Symptom/25 points	Satisfaction/40 points	Expectation/15 points	Activity/100 points
Pain/20 points	11.63 ± 4.25	−0.53* (< 0.001)	−0.50* (< 0.001)	0.13 (0.167)	−0.50* (< 0.001)
Stiffness/8 points	4.71 ± 2.13	−0.35* (< 0.001)	−0.47* (< 0.001)	0.15 (0.095)	−0.38* (< 0.001)
Function/68 points	40.21 ± 14.62	−0.51* (< 0.001)	−0.46* (< 0.001)	0.23 (0.011)*	−0.54* (< 0.001)
Total/96 points	56.55 ± 19.90	−0.53* (< 0.001)	−0.49* (< 0.001)	0.21 (0.019)*	−0.53* (< 0.001)

KSS = Knee Society Score; Spearman's correlation coefficients (r) when comparing the four KSS subscales with the WOMAC (p value); *significant correlation at $p < 0.05$.

Table 4. Construct validity between the Korean New KSS and the Korean SF-36

Korean SF-36		Korean New KSS (preoperative form)			
Subscales	Score Mean ± SD	Symptom/25 points	Satisfaction/40 points	Expectation/15 points	Activity/100 points
Physical function	25.93 ± 7.37	0.41* (< 0.001)	0.32* (< 0.001)	−0.22* (0.017)	0.62* (< 0.001)
Role-physical	30.85 ± 10.84	0.14 (0.116)	0.23* (0.012)	−0.11 (0.208)	0.46* (< 0.001)
Bodily pain	33.84 ± 10.43	0.22* (0.016)	0.33* (< 0.001)	−0.11 (0.225)	0.26* (0.003)
General health	40.64 ± 8.41	−0.04 (0.629)	0.12 (0.201)	0.21* (0.017)	0.05 (0.625)
Vitality	43.51 ± 10.07	0.11 (0.218)	0.23* (0.012)	−0.01 (0.876)	0.15 (0.107)
Social function	40.25 ± 13.08	0.20* (0.025)	0.28* (0.002)	−0.04 (0.631)	0.31* (0.001)
Role-emotion	35.52 ± 16.06	0.09 (0.323)	0.21* (0.022)	0.06 (0.545)	0.18 (0.052)
Mental health	44.70 ± 10.16	0.09 (0.299)	0.21* (0.022)	−0.03 (0.782)	0.06 (0.491)
Physical component summary	29.41 ± 7.77	0.26* (0.003)	0.29* (0.001)	−0.14 (0.134)	0.52* (< 0.001)
Mental component summary	46.58 ± 12.75	0.08 (0.382)	0.21* (0.022)	0.07 (0.454)	0.10 (0.280)

KSS = Knee Society Score; Spearman's correlation coefficients (r) when comparing the four KSS subscales with the Korean SF-36 (p value); *significant correlation at $p < 0.05$.

activity domains to all or most of the subscales of SF-36 (Table 4). The satisfaction domain showed a weak positive correlation with all the subscales of the Korean SF-36 except general health (r = 0.32, p < 0.001; r = 0.23, p = 0.012; r = 0.33, p < 0.001; r = 0.23, p = 0.012; r = 0.28, p = 0.002; r = 0.21, p = 0.022; r = 0.21, p = 0.022; r = 0.29, p = 0.001; and r = 0.21, p = 0.022, respectively). The activity domain showed a strong positive correlation with physical function (r = 0.62, p < 0.001) and physical component summary (r = 0.52, p < 0.001), moderate with physical role (r = 0.46, p < 0.001), and weak with bodily pain (r = 0.26, p = 0.003) and social function (r = 0.31, p = 0.001). The symptom domain also exhibited a similar moderate positive correlation with physical function (r = 0.41, p < 0.001) and weak with bodily pain (r = 0.22, p = 0.016), social function (r = 0.20, p = 0.025), and physical component summary (r = 0.26, p = 0.003). The expectation domain showed a weak negative correlation with physical function (r = -0.22, p = 0.017) and weak positive correlation with general health (r = 0.21, p = 0.017).

The Korean New KSS was found to be more responsive than the Korean WOMAC and Korean SF-36. All the domains of the Korean New KSS, except for expectation, showed a large change (> 0.8), calculated as SRM. This analysis showed that the Korean New KSS having a SRM of 2.03 (p < 0.001) was more responsive than the Korean WOMAC with a SRM of 1.88 (p < 0.001) and the Korean SF-36 physical and mental component summaries, with SRMs of 1.14 (p < 0.001) and 0.68 (p < 0.001) respectively. The SRM of the Korean KSS symptom score was 2.23 (p < 0.001) which was higher than the Korean WOMAC pain (2.12, p < 0.001) and SF-36 bodily pain scores (1.14, p < 0.001). Furthermore, regarding the

functional scale, the Korean New KSS had a mean score of 1.85 (p < 0.001) which indicates that it was more responsive than the Korean WOMAC with a score of 1.75 (p < 0.001) and the SF-36 physical function score of 1.67 (p < 0.001) (Table 5).

Construct validity and responsiveness for age-matched control groups between the Korean New KSS, Korean WOMAC, and Korean SF-36 also showed good correlations. Gender analysis between the Korean New KSS and Korean WOMAC showed strong to moderate correlations with the exception of the expectation subscale which correlated weakly (Table 6). Similarly, the Korean New KSS and Korean SF-36 showed moderate to weak positive correlation when men and women were compared individually (Table 7). Furthermore, gender analysis of responsiveness showed that the Korean New KSS had higher SRMs when compared with the Korean WOMAC and Korean SF-36 (Table 8).

Discussion

The recently developed 2011 KS Score is widely used for patients undergoing TKA. For a measure to be effective across cultures, it not only has to be translated well linguistically, but also has to be adapted culturally to maintain the content validity of the instrument [8]. The current study was conducted to develop the Korean version of the 2011 KS Score for Korean-speaking patients who undergo TKA, which was developed through crosscultural adaptation, and to investigate its psychometric properties. Similar versions of the 2011 KS Score have been translated to Japanese [10], French [6], Dutch [22], and Chinese [16]

Table 5. Responsiveness of the Korean New KSS compared with the Korean WOMAC and Korean SF-36

Questionnaire	Mean of change	SD	SRM* (95% CI)	p Value
Korean New KSS				
Symptom (3 items)/25 points	14.88	6.69	2.23 (1.83–2.65)	< 0.001
Satisfaction (5 items)/40 points	15.33	8.73	1.76 (1.49–2.01)	< 0.001
Expectation (3 items)/15 points	−1.45	3.69	−0.39 (−0.58 to −0.20)	< 0.001
Activity (19 items)/100 points	40.54	21.86	1.85 (1.33–2.19)	< 0.001
Amount (30)/180 points	69.30	34.07	2.03 (1.63–2.37)	< 0.001
Korean WOMAC				
Pain/20 points	−10.02	4.73	−2.12 (−2.42 to −1.80)	< 0.001
Stiffness/8 points	−3.07	2.72	−1.13 (−1.39 to −0.87)	< 0.001
Function/68 points	−29.21	16.69	−1.75 (−1.98 to −1.53)	< 0.001
Total/96 points	−42.31	22.55	−1.88 (−2.12 to −1.64)	< 0.001
Korean SF-36				
Physical function	16.05	9.61	1.67 (1.42–1.90)	< 0.001
Role-physical	13.48	16.29	0.83 (0.60–1.05)	< 0.001
Bodily pain	14.80	12.97	1.14 (0.85–1.36)	< 0.001
General health	4.46	26.05	0.17 (−0.06 to 0.25)	0.030
Vitality	8.07	11.91	0.68 (0.4–0.84)	< 0.001
Social function	12.04	15.17	0.79 (0.60–1.00)	< 0.001
Role-emotion	16.81	17.14	0.98 (0.79–1.20)	< 0.001
Mental health	7.46	11.76	0.63 (0.47–0.81)	< 0.001
Physical component summary	12.37	10.82	1.14 (0.93–1.37)	< 0.001
Mental component summary	9.15	13.33	0.68 (0.51–0.85)	< 0.001

KSS = Knee Society Score; *calculated as the mean change between the preoperative and 12-month scores divided by the SD of the change in score; standardized response mean (SRM) of 0.2–0.5 = small change, 0.5–0.8 = moderate change, and ≥ 0.8 = large change; paired t-test, $p < 0.05$.

Table 6. Construct validity between the Korean New KSS and the Korean WOMAC for men and women

Korean WOMAC Subscales	Score Mean ± SD	Korean New KSS (preoperative form) Symptom/25 points	Satisfaction/40 points	Expectation/15 points	Activity/100 points
Men (n = 53)					
Pain/20 points	9.49 ± 4.53	−0.59* (< 0.001)	−0.69* (< 0.001)	0.01 (0.970)	−0.61* (< 0.001)
Stiffness/8 points	4.08 ± 1.95	−0.51* (< 0.001)	−0.61* (< 0.001)	0.11 (0.446)	−0.48* (< 0.001)
Function/68 points	29.02 ± 12.92	−0.70* (< 0.001)	−0.73* (< 0.001)	0.21* (0.026)	−0.78* (< 0.001)
Total/96 points	42.58 ± 18.14	−0.71* (< 0.001)	−0.76* (< 0.001)	0.20* (0.041)	−0.76* (< 0.001)
Women (n = 106)					
Pain/20 points	11.59 ± 4.32	−0.55* (< 0.001)	−0.51* (< 0.001)	0.17 (0.083)	−0.52* (< 0.001)
Stiffness/8 points	4.72 ± 2.18	−0.32* (0.001)	−0.47* (< 0.001)	0.19 (0.050)	−0.39* (< 0.001)
Function/68 points	40.74 ± 14.89	−0.52* (< 0.001)	−0.46* (< 0.001)	0.24* (0.011)	−0.54* (< 0.001)
Total/96 points	57.05 ± 20.29	−0.54* (< 0.001)	−0.49* (< 0.001)	0.23* (0.015)	−0.54* (< 0.001)

KSS = Knee Society Score; Spearman's correlation coefficients (r) when comparing the four Korean KSS subscales with the Korean WOMAC (p value); *significant correlation at $p < 0.05$.

through the process of crosscultural adaptation, and all have been shown to have good psychometric properties. The current study similarly shows that the Korean New KSS is a reliable, consistent, and valid instrument to evaluate the functional outcomes and expectations of Korean-speaking patients before and after TKA (Appendix

Table 7. Construct validity between the Korean New KSS and the Korean SF-36 for men and women

Korean SF-36 Subscales	Score Mean ± SD	Korean New KSS (preoperative form)			
		Symptom/25 points	Satisfaction/40 points	Expectation/15 points	Activity/100 points
Men (n = 53)					
Physical function	1.92 ± 8.12	0.44* (0.001)	0.44* (0.001)	−0.34* (0.013)	0.51* (< 0.001)
Role-physical	37.39 ± 10.08	0.40* (0.003)	0.28* (0.041)	−0.17 (0.227)	0.37* (0.006)
Bodily pain	34.66 ± 8.59	0.57* (< 0.001)	0.58* (< 0.001)	−0.15 (0.292)	0.43* (0.001)
General health	43.31 ± 9.35	0.33* (0.016)	0.33* (0.015)	0.19 (0.181)	0.31* (0.026)
Vitality	48.53 ± 8.91	0.41* (0.003)	0.38* (0.006)	0.14 (0.311)	0.31* (0.022)
Social function	42.01 ± 12.19	0.49* (< 0.001)	0.34* (0.012)	−0.05 (0.716)	0.40* (0.003)
Role-emotion	39.13 ± 14.79	0.38* (0.006)	0.33* (0.015)	−0.01 (0.953)	0.32* (0.020)
Mental health	48.69 ± 9.97	0.41* (0.003)	0.30* (0.027)	0.28* (0.043)	0.39* (0.004)
Physical component summary	33.75 ± 7.34	0.42* (0.002)	0.38* (0.005)	−0.26 (0.058)	0.44* (0.001)
Mental component summary	49.47 ± 10.94	0.47* (< 0.001)	0.39* (0.004)	0.20 (0.147)	0.42* (0.002)
Women (n = 106)					
Physical function	25.71 ± 7.51	0.43* (< 0.001)	0.30* (0.002)	−0.16 (0.092)	0.63* (< 0.001)
Role-physical	30.64 ± 11.08	0.16 (0.103)	0.22* (0.021)	−0.08 (0.387)	0.46* (< 0.001)
Bodily pain	34.08 ± 10.70	0.20* (0.032)	0.31* (0.001)	−0.13 (0.176)	0.27* (0.005)
General health	40.39 ± 8.34	−0.07 (0.493)	0.09 (0.333)	0.23* (0.017)	0.03 (0.762)
Vitality	43.09 ± 10.01	0.11 (0.250)	0.21* (0.027)	−0.05 (0.597)	0.15 (0.125)
Social function	40.27 ± 13.10	0.20* (0.037)	0.26* (0.005)	−0.01 (0.904)	0.27* (0.004)
Role-emotion	35.80 ± 16.22	0.09 (0.340)	0.19* (0.046)	0.06 (0.538)	0.20* (0.034)
Mental health	44.53 ± 10.24	0.85 (0.374)	0.20* (0.036)	−0.08 (0.379)	0.05 (0.586)
Physical component summary	29.23 ± 7.94	0.27* (0.004)	0.28* (0.003)	−0.09 (0.350)	0.51* (< .001)
Mental component summary	46.58 ± 13.07	0.07 (0.486)	0.18 (0.060)	0.05 (0.629)	0.09 (0.369)

KSS = Knee Society Score; Spearman's correlation coefficients (r) when comparing the four Korean KSS subscales with the Korean SF-36 (p value); *significant correlation at p < 0.05.

1. Supplemental material is available with the online version of *CORR*®.).

We acknowledge certain limitations of our study. First, because all three questionnaires were administered simultaneously 4 weeks preoperatively and 1 year postoperatively, the patients could have been affected by responder burden, which might lead to similar or missing responses. This was addressed by having a trained research assistant review the questionnaires for such responses in the presence of eligible patients without any attempt to influence the response. However, we acknowledge that by having research personnel assist in the administration of what is designed as a patient-reported outcomes tool, the performance with the Korean version of the KSS with respect to reliability and validity we observed in this study may be somewhat better than might be achieved with unassisted patients in general use. Second, the demographic features of TKA use in Korea, such as the predominance of older women, should be taken into account; further studies are needed to investigate these psychometric properties in Korean men, although we attempted to address that to some extent with our additional analysis. Third, although our translation methods were rigorous, certain inconsistencies might remain in the translation from one language to another. If better words or phrases are suggested, they should undergo validation by using the same standardized protocol.

The test-retest reliability was excellent for all the domains of the Korean New KSS, showing a high degree of concordance (ICC) and good to excellent internal consistency. Reliability refers to the degree to which test results have a brief internal consistency without any treatment change, and internal consistency is defined as the coherence among scale components. Some authors quote a period between 2 days and 2 weeks for the second application of a test, which is an adequate compromise between recall bias and change in disease condition [17]. However, according to Terwee et al. [21], the appropriateness of the period chosen is not as important as the justification of the period described. In the current study, reliability and internal consistency were assessed by asking patients to complete the questionnaires twice during a 4-week interval.

Table 8. Responsiveness of the Korean New KSS compared with the Korean WOMAC and Korean SF-36

Questionnaire	Mean of change	SD	SRM* (95% CI)	p Value
Korean New KSS				
Men (n = 53)				
Symptoms (3 items)/25 points	13.74	5.27	2.61 (2.12–3.20)	< 0.001
Satisfaction score (5 items)/40 points	19.83	11.93	1.66 (1.21–2.06)	< 0.001
Expectation score (3 items)/15 points	−1.34	3.73	−0.36 (−0.60 to −0.10)	0.006
Activity (19 items)/100 points	32.70	21.90	1.49 (1.12–1.87)	< 0.001
Amount (30 items)/180 points	64.92	29.05	2.24 (1.80–2.65)	< 0.001
Korean WOMAC				
Pain/20 points	−8.43	4.23	−1.99 (−2.45 to −1.60)	< 0.001
Stiffness/8 points	−.2.45	2.26	−1.08 (−1.41 to −0.81)	< 0.001
Function/68 points	−20.28	13.70	−1.48 (−1.91 to −1.00)	< 0.001
Total/96 points	−31.17	18.29	−1.70 (−2.12 to −1.22)	< 0.001
Korean SF-36				
Physical function	13.49	9.32	1.45 (1.06–1.78)	< 0.001
Role-physical	11.10	13.39	0.83 (0.53–1.17)	< 0.001
Bodily pain	17.38	11.16	1.56 (1.19–1.98)	< 0.001
General health	4.17	8.31	0.50 (0.20–0.76)	< 0.001
Vitality	5.83	8.24	0.71 (0.39–0.95)	< 0.001
Social function	11.11	13.10	0.84 (0.54–1.12)	< 0.001
Role-emotion	11.58	15.15	0.76 (0.51–1.01)	< 0.001
Mental health	3.55	10.30	0.34 (0.05–0.61)	0.008
Physical component summary	13.26	9.44	1.41 (1.05–1.81)	< 0.001
Mental component summary	4.90	10.39	0.47 (0.22–0.70)	< 0.001
Korean New KSS				
Women (n = 106)				
Symptoms (3 items)/25 points	14.95	6.91	2.16 (1.70–2.59)	< 0.001
Satisfaction score (5 items)/40 points	15.10	8.90	1.70 (1.44–1.97)	< 0.001
Expectation score (3 items)/15 points	−1.59	3.71	−0.43 (−0.63 to −0.22)	< 0.001
Activity (19 items)/100 points	40.50	22.29	1.82 (1.39–2.18)	< 0.001
Amount (30 items)/180 points	68.95	34.70	1.99 (1.60–2.35)	< 0.001
Korean WOMAC				
Pain/20 points	−9.99	4.85	−2.06 (−2.36 to −1.76)	< 0.001
Stiffness/8 points	−3.04	2.81	−1.08 (−1.33 to −0.81)	< 0.001
Function/68 points	−29.50	17.04	−1.73 (−1.96 to −1.50)	< 0.001
Total/96 points	−42.53	23.12	−1.84 (−2.07 to −1.60)	< 0.001
Korean SF-36				
Physical function	16.22	9.70	1.67 (1.44–1.94)	< 0.001
Role-physical	14.83	16.44	0.90 (0.66–1.15)	< 0.001
Bodily pain	14.76	13.20	1.12 (0.80–1.35)	< 0.001
General health	4.57	27.36	0.17 (−0.04 to 0.24)	0.040
Vitality	8.18	12.12	0.68 (0.47–0.84)	< 0.001
Social function	11.96	15.14	0.79 (0.59–1.00)	< 0.001
Role-emotion	16.80	17.38	0.97 (0.75–1.17)	< 0.001
Mental health	7.69	11.92	0.65 (0.47–0.80)	< 0.001
Physical component summary	12.52	10.92	1.15 (0.91–1.39)	< 0.001
Mental component summary	9.18	13.76	0.67 (0.48–0.85)	< 0.001

KSS = Knee Society Score; *calculated as the mean change between the preoperative and 12-month scores divided by the SD of the change in score; standardized response mean (SRM) of 0.2–0.5 = small change, 0.5–0.8 = moderate change, and ≥ 0.8 = large change; paired t-test, p < 0.05.

Table 9. Comparison of test-retest reliability and internal consistency

KSS domain	Korean		Dutch		Japanese		French	
	ICC	CA	ICC	CA	ICC	CA	ICC	CA
Symptoms (3 items)/25 points	0.85	0.92	0.92	0.96	0.65	0.78	0.97	0.70
Satisfaction score (5 items)/40 points	0.81	0.89	0.73	0.84	0.88	0.94	0.84	0.90
Expectation score (3 items)/15 points	0.80	0.89	0.84	0.91	0.83	0.91	0.87	0.80
Activity (19 items)/100 points	0.81	0.90	0.87	0.93	NA	NA	0.94	0.80
Total score/180 points	0.86	0.93	0.82	0.93	NA	NA	0.96	0.80

KSS = Knee Society Score; ICC = intraclass correlation coefficient; CA = Cronbach's alpha; NA = not available.

The Korean New KSS showed excellent reliability for all domains (ICC, 0.69–0.85), showing good reproducibility. It also showed good-to-excellent internal consistency in all the subscales (Cronbach's alpha, 0.83–0.92). Our study showed an ICC of 0.69 to 0.85, which is similar to the results of Japanese [10], French [6], and Dutch [22] studies that translated the 2011 KS Score to their respective languages, with ICCs ranging from 0.65 to 0.88, from 0.84 to 0.97, and from 0.73 to 0.92, respectively. Similarly, our study showed a Cronbach's alpha of 0.83 to 0.92, which also is in agreement with those of other translated versions of the 2011 KS Score and the original English KSS (Cronbach's alpha, 0.68–0.95) [19], showing g good-to-excellent reliability and internal consistency (Table 9). Although we could not provide comparative data regarding reliability for age-matched control groups as reliability in our study was assessed at an interval 4 weeks apart, our data regarding validity and responsiveness appear adequate to address the gender disparity of our study.

The Korean New KSS showed adequate construct validity when compared with the Korean WOMAC and Korean SF-36. Because no gold standard measure has been established to evaluate validity post-TKA, correlations between the preoperative scores of the Korean New KSS and those of the Korean WOMAC and Korean SF-36 were determined. These scoring systems were selected because their test validities have been proven in previous studies [2, 11]. All the domains of the Korean New KSS correlated well with the Korean WOMAC, except for the expectation subscale. All the domains tested showed strong or moderate correlation with the individual subscales of the Korean WOMAC. The satisfaction domain of the Korean New KSS showed a weak positive correlation with all the subscales of the SF-36 except general health, which might be expected owing to the high satisfaction rates post-TKA. Similarly, some studies have indicated that the current TKA population is physically more active than in the past and observed that some patients start participating in physical activities postoperatively, which they were not able to do preoperatively [1, 12]. Therefore, the activity domain becomes an important tool for post-TKA evaluation. In the current study, the activity domain of the Korean New KSS showed a strong positive correlation with physical function (r = 0.62, p < 0.001) and physical component summary (r = 0.52, p < 0.001), moderate with physical role (r = 0.46, p < 0.001), and weak with bodily pain (r = 0.26, p = 0.003) and the social function (r = 0.31, p = 0.001) component of the Korean SF-36. Similarly, the symptom domain also exhibited similar moderate positive correlation with physical function (r = 0.41, p < 0.001) and weak positive correlation with bodily pain, social function, and the physical component summary (r = 0.22, p = 0.016; r = 0.20, p = 0.025; and r = 0.26, p = 0.003, respectively). The expectation domain showed a similar correlation with the physical function and general health subscales of the Korean SF-36. Our study shows low levels of correlation when compared with the Dutch [22] and French studies [6] but similar correlations when compared with the Japanese version of the 2011 KS Score [10]. A possible explanation for this finding could be the difference in cultural background between the European and Asian populations. Nevertheless, such differences in correlation coefficients do not reduce the usefulness of our study but rather indicate that further studies are needed to identify the reason for these findings and their effect on the instrument. Our data regarding age-matched control groups showed similar values for Korean men and women when construct validity was analyzed individually, thereby eliminating any gender-based biases from our study.

The Korean New KSS was the most responsive scale when compared with the Korean WOMAC and Korean SF-36. Responsiveness shows the ability of a scale to reflect changes in perioperative and postoperative results; the higher the responsiveness, the greater the ability of a scale to detect changes [3]. Analysis also revealed that the Korean New KSS symptom score was more responsive when compared with the WOMAC pain and the SF-36 bodily pain scores. However, regarding the functional scale, the Korean New KSS was more responsive than the Korean WOMAC and the Korean SF-36. Similarly, for

age-matched control groups, the Korean New KSS also was more responsive when compared with the Korean WOMAC and Korean SF-36, thereby confirming our current study population is representative of the Korean population undergoing TKA.

The Korean version of the 2011 KS Score appears valid, reliable, and responsive in Korean-speaking patients who undergo TKA for knee osteoarthritis. Therefore, it now can be used as a valuable metric to assess functional outcomes and expectations of Korean patients who undergo TKA. Because the population of men undergoing knee arthroplasty in Korea is small compared with that of women, further studies will be required to investigate the properties of the Korean New KSS among men.

References

1. Argenson JN, Parratte S, Ashour A, Komistek RD, Scuderi GR. Patient-reported outcome correlates with knee function after a single-design mobile-bearing TKA. *Clin Orthop Relat Res.* 2008;466:2669–2676.

2. Bae SC, Lee HS, Yun HR, Kim TH, Yoo DH, Kim SY. Cross-cultural adaptation and validation of Korean Western Ontario and McMaster Universities (WOMAC) and Lequesne osteoarthritis indices for clinical research. *Osteoarthritis Cartilage.* 2001;9:746–750.

3. Beck CT, Gable RK. Ensuring content validity: an illustration of the process. *J Nurs Meas.* 2001;9:201–215.

4. Bland JM, Altman DG. Measurement error. *BMJ.* 1996;312:1654.

5. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken).* 2011;63(suppl 11):S208–228.

6. Debette C, Parratte S, Maucort-Boulch D, Blanc G, Pauly V, Lustig S, Servien E, Neyret P, Argenson JN. French adaptation of the new Knee Society Scoring System for total knee arthroplasty. *Orthop Traumatol Surg Res.* 2014;100:531–534.

7. Ghanem E, Pawasarat I, Lindsay A, May L, Azzam K, Joshi A, Parvizi J. Limitations of the Knee Society Score in evaluating outcomes following revision total knee arthroplasty. *J Bone Joint Surg Am.* 2010;92:2445–2451.

8. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol.* 1995;24:61–63.

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;46:1417–1432.

10. Hamamoto Y, Ito H, Furu M, Ishikawa M, Azukizawa M, Kuriyama S, Nakamura S, Matsuda S. Cross-cultural adaptation and validation of the Japanese version of the new Knee Society Scoring System for osteoarthritic knee with total knee arthroplasty. *J Orthop Sci.* 2015;20:849–853.

11. Han CW, Lee EJ, Iwaya T, Kataoka H, Kohzuki M. Development of the Korean version of Short-Form 36-Item Health Survey: health related QOL of healthy elderly people and elderly patients in Korea. *Tohoku J Exp Med.* 2004;203:189–194.

12. Huch K, Muller KA, Sturmer T, Brenner H, Puhl W, Gunther KP. Sports activities 5 years after total knee or hip arthroplasty: the Ulm Osteoarthritis Study. *Ann Rheum Dis.* 2005;64:1715–1720.

13. Insall J. Current concepts review: patellar pain. *J Bone Joint Surg Am.* 1982;64:147–152.

14. Koh IJ, Kim TK, Chang CB, Cho HJ, In Y. Trends in use of total knee arthroplasty in Korea from 2001 to 2010. *Clin Orthop Relat Res.* 2013;471:1441–1450.

15. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28:632–642.

16. Liu D, He X, Zheng W, Zhang Y, Li D, Wang W, Li J, Xu W. Translation and validation of the simplified Chinese new Knee Society Scoring System. *BMC Musculoskelet Disord.* 2015;16:391.

17. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol.* 2003;56:730–735.

18. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4:10.

19. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, Chadha P, Daylamani DA, Scott WN, Bourne RB. Development of a new Knee Society scoring system. *Clin Orthop Relat Res.* 2012;470:20–32.

20. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The new Knee Society Knee Scoring System. *Clin Orthop Relat Res.* 2012;470:3–19.

21. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.

22. Van Der Straeten C, Witvrouw E, Willems T, Bellemans J, Victor J. Translation and validation of the Dutch new Knee Society Scoring System©. *Clin Orthop Relat Res.* 2013;471:3565–3571.