

## Translation and Validation of the Dutch New Knee Society Scoring System<sup>©</sup>

Catherine Van Der Straeten MD, Erik Witvrouw PT, PhD,  
Tine Willems PT, PhD, Johan Bellemans MD, PhD,  
Jan Victor MD, PhD

Received: 11 January 2013 / Accepted: 25 June 2013 / Published online: 9 July 2013  
© The Association of Bone and Joint Surgeons® 2013

### Abstract

**Background** A new version of The Knee Society Knee Scoring System<sup>©</sup> (KSS) has recently been developed. Before this scale can be used in non-English-speaking populations, it has to be translated and validated for a particular population.

**Questions/purposes** We evaluated the construct and content validity, the test-retest reliability, and the internal consistency of the Dutch version of the New Knee Society KSS.

**Methods** A Dutch translation was performed using a forward-backward translation protocol. We tested the construct

validity of the Dutch New KSS by comparing it with the Dutch versions of the WOMAC, Knee Injury and Osteoarthritis Outcome Score (KOOS), and SF-12 scores in 137 patients undergoing total knee arthroplasty (TKA). Content validity was assessed by comparing pre- and postoperative scores and by checking floor and ceiling effects. To evaluate test-retest reliability and consistency, 47 patients completed the questionnaire a second time with a mean of 8 days interval (range, 2–20 days) between tests.

**Results** Construct validity was demonstrated because the Dutch New KSS correlated well with the Dutch WOMAC ( $r = -0.751$ ;  $p < 0.001$ ), Dutch KOOS ( $r = -0.723$ ;  $p < 0.001$ ), and Dutch SF-12 ( $r = 0.569$ ;  $p < 0.001$ ). There was a significant difference between pre- and postoperative scores ( $p < 0.001$ ) in line with the other scores. Test-retest reliability proved excellent with an intraclass correlation coefficient between 0.73 and 0.92 depending on the domain tested. Consistency as indicated by Cronbach's alpha ranging from 0.84 to 0.96 was good to excellent.

**Conclusions** As demonstrated by the validation procedure, the Dutch New KSS is an excellent instrument to evaluate TKA outcome in Dutch-speaking patients.

**Level of Evidence** Level II, diagnostic study. See Guidelines for Authors for a complete description of levels of evidence.

---

Each author certifies that he or she, or a member of his or her immediate family, has no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.

Each author certifies that his or her institution approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

This work was performed at Ghent University Hospital, Ghent, Belgium, and the University Hospital of Leuven, Leuven, Belgium.

---

C. Van Der Straeten (✉), T. Willems, J. Victor  
Department of Orthopedic Surgery and Physical Medicine,  
Faculty of Medicine & Health Sciences, Ghent University,  
University Hospital, De Pintelaan 185, 9000 Ghent, Belgium  
e-mail: Catherine.VanDerStraeten@ugent.be

J. Bellemans  
Department of Orthopaedic Surgery and Traumatology,  
University Hospitals Leuven, Catholic University Leuven,  
University Hospital, Pellenberg, Belgium

E. Witvrouw  
Department of Rehabilitation Sciences, Faculty of Medicine &  
Health Sciences, Ghent University, University Hospital, Ghent,  
Belgium

## Introduction

Orthopaedic outcome measurements have traditionally focused on objective parameters such as radiographic analysis or clinical tests. In the last decades, the development of validated patient-oriented measures through questionnaires has added another dimension to clinical outcome evaluation [7, 9, 27]. Health-related quality of life can be considered as the perception of the patient's health and it is a fundamental outcome measure for clinical research in orthopaedics. In 1989, The Knee Society Clinical Rating System was developed as a simple, objective scoring system to rate the knee and patient's functional abilities such as walking and stairclimbing before and after TKA [16]. The Knee Society Clinical Rating System has been the most popular method of tracking and reporting outcomes after total and partial knee arthroplasty worldwide [24]. However, it became clear that certain ambiguities and deficiencies within the original Knee Society Clinical Rating System challenged its use and validity [11]. In the last decades, the number of young and active patients undergoing TKA has rapidly increased. These patients are expected to live longer after TKA, sustain a higher functional and recreational activity level, and are less willing to accept physical limitations. A number of frequently used patient self-reported outcome questionnaires such as the WOMAC [3] and the SF-36 or SF-12 are considered reference scores [5, 19, 20] for the assessment of patient-relevant treatment effects in osteoarthritis (OA) but are not specifically responsive to knee treatment. The Knee Injury and Osteoarthritis Outcome Score (KOOS) [23] was developed to cover the expectations of younger and more active patients with a knee injury or knee OA regarding more demanding physical functions but is not considered the gold standard score because of its weak validity regarding the sport and recreation subscale [5]. Therefore, a new Knee Society Knee Scoring System (New KSS) has recently been developed [21] to document the expectations, satisfaction, and physical activities of current patients undergoing TKA more accurately. The score was validated in a methodological fashion confirming the reliability and the consistency of the total new KSS and its different domains (Objective Knee Score including a Pain Score, Satisfaction Score, Expectation Score, and Functional Activity Score subdivided in four subscales of walking and standing, standard activities, advanced activities, and discretionary activities), which were also analyzed for differential item functioning [21].

Questionnaires designed as subjective scoring systems can be used in countries other than the ones in which they were developed if they are translated and validated for a target language population. Because the former Knee Society Clinical Rating System was a very popular score in

The Netherlands and in Flanders (Dutch-speaking part of Belgium), there was a strong need to translate and validate a Dutch version of the New KSS. Besides, this Dutch version is imperative for the setup of international multicenter studies. Therefore, the purpose of the study was to validate the Dutch translation of the New KSS for Dutch-speaking patients undergoing TKA. The specific aims of the research were to investigate (1) the construct validity of the Dutch New KSS, defined as its quality to measure the intended construct, ie, the response to knee treatment, in comparison with other scores with proven validity such as WOMAC, KOOS, and SF-12; (2) the content validity of the Dutch New KSS, which is a nonstatistical validity test focused on determining whether the content of the test represents its intended objectives; (3) the test-retest reliability, ie, the ability of the score to reproduce similar results on different occasions separated by a relatively brief time interval; and (4) the internal consistency of the Dutch New KSS, which refers to the coherence of the different components of the scale.

## Patients and Methods

### Translation Procedure

Dutch translation was made using a forward-backward translation protocol according to the guidelines of Guillemin et al. [12, 13]. An English version of the New Knee Society Score was translated into Dutch by two independent translators (JB, JV). Discrepancies between the two versions were resolved through consensus. An initial Dutch version was produced. This version was translated back into English by two independent translators (JB, JV) not involved in the initial steps. These versions were compared with the original by a committee, and the final version was made. Finally, this version was pretested on 20 patients. Comments on clarities of wordings and problems encountered during the self-administration process were collected. Because the questions of the New KSS address problems universally encountered by all patients undergoing knee arthroplasty and because there are no major cultural differences among Flemish, Dutch, and American populations regarding activities or expectations [22], it was not deemed necessary to make cultural adaptations to the questionnaire.

### Patients

A total of 137 patients participated in this study (Table 1). All patients had OA and were candidates for TKA (n = 48) or had undergone a TKA (n = 89) and presented for their 6-month followup. Half of these patients were

**Table 1.** Anthropometric data of the total patient population (n = 137) and the test-retest patient population (n = 47)

Variable		Total population		Test-retest patient population	
		N = 137	Percent	N = 47	Percent
Sex	Male	61	44.5	22	47
	Female	76	55.5	25	53
Preoperative		48	35	40	85
Postoperative		89	65	7	15
Side	Left	79	58	24	51
	Right	58	42	23	49
Length (cm)	Mean	167.6		168.3	
	(SD)	(10.6)		(10.7)	
Weight (kg)	Mean	82.2 (16.6)		83.5 (18.1)	
	(SD)				
Age (years)	Mean	63.5 (12.6)		57.5 (13.8)	
	(SD)				

recruited consecutively from the Ghent University Hospital, Ghent, Belgium, and the other half from the University Hospital in Leuven, Belgium. The total sample size of 137 patients was similar to the number of 145 used for the validation of the Dutch International Knee Documentation Committee Subjective Knee Form [15]. All participants were properly informed about the study goals and the setup and signed an informed consent form. The study was approved by both local ethics committees.

### Questionnaires

All patients were asked to answer a questionnaire containing the translated New Knee Society KSS besides the Dutch WOMAC score, the Dutch KOOS, and the SF-12 [1, 3, 14, 15, 17].

Validity of a questionnaire refers to comparison to a gold standard test, which is not available for TKA to date [5]. Consequently, the validity of the construct of the Dutch New KSS with regard to its ability to measure what it was intended for (ie, the effect of TKA) was evaluated by analyzing the relationships between responses to different test items and the relationships between the test and other constructs (scores). We tested the construct validity of the Dutch New KSS by comparing it with the Dutch WOMAC, Dutch KOOS, and the Dutch SF-12. The validity domains of the Dutch New KSS were tested against the appropriate subscales of the other scores. WOMAC and KOOS scores were chosen because they are frequently used to assess OA in general (WOMAC) and patients with knee disorders in particular (KOOS) and have proven internal consistency and test-retest reliability for patients with knee disorders

[3, 5]. The KOOS consists of five subscales (pain, symptoms, activities of daily living, sport and recreation function, and knee-related quality of life) each with an excellent internal consistency [5, 8, 23]. The SF-12 is a brief, 12-item version of the widely used SF-36 [19, 20, 25, 26]. The SF-12 identifies two summary scales, termed the Physical Component Summary scale (PCS) and the Mental Component Summary scale (MCS), and consistently differentiates presence and seriousness of physical and mental conditions between groups with adequate test-retest reliabilities [1].

To determine the test-retest reliability and the internal consistency of the Dutch New KSS, a random sample of 53 patients were asked to fill out the questionnaire a second time under similar circumstances with a 1-week interval and return them immediately after completion. The 1-week interval was chosen as a reference interval because Marx et al. [18] could not find a significant difference between repeated tests at 2 days or 2 weeks [14, 15]. The sample size needed was calculated based on the population and reliability results of the Dutch translation of the Oxford Score [14, 15]. For a Type I error rate, alpha of 0.05, and a power of 0.90, the sample size needed was 38; for a power of 0.95, the sample size needed was 48. Hence, taking into account a possible 10% loss to followup, 53 patients were asked to complete the forms twice.

### Statistical Analysis

Construct validity of was evaluated by calculation of the Pearson correlation coefficients as a measure of the strength of the correlation between the scores of the different scales and subscales. A correlation  $r > 0.5$  (absolute value) was considered strong. The construct validity of the different domains of the Dutch New KSS were also tested against the KOOS subscales.

Convergent construct validity of the Dutch New KSS was tested by correlating it to the WOMAC, KOOS, and PCS of the SF-12 scores, both for preoperative and postoperative questionnaires. Divergent or discriminant construct validity was assessed by correlating the Dutch New KSS results with the MCS of the SF-12. It was hypothesized that correlations SF-12 physical subscales would be strong ( $r > 0.5$ ; convergent validity) but weak with the SF-12 mental subscales ( $r < 0.5$ ; divergent validity) [14].

Content validity of a questionnaire is a nonstatistical term and relies more on expert opinion to determine whether the scale covers a representative sample of the domain to be measured [2, 10]. Content validity can be examined by evaluating the responsiveness of the Dutch New KSS to the treatment (TKA). Preoperative and

postoperative scores were compared with nonparametric statistical tests (Mann-Whitney U, Kruskal-Wallis) and also assessed against WOMAC, KOOS, and SF-12 results (paired t-tests). The occurrence of floor effects (large number of subjects with poorest score of 0) and ceiling effects (large number of subjects reaching best score of 100) was also examined to evaluate the ability of the Dutch New KSS to distinguish further improvements and its responsiveness to smaller effect changes. These effects exist when a questionnaire repetitively scores the maximum or minimum value and represent a measuring limitation because they indicate a questionnaire may not be set properly to what it is intended to measure.

Test-retest reliability was assessed using the intraclass correlation coefficient (ICC; Pearson correlation) [4] between the first and second applications of the Dutch New KSS. Paired t-tests were performed to determine the systematic difference between the first and second tests. Internal consistency was assessed using Cronbach's alpha [6], which is a complementary measure of reliability by determining the average correlation of items in a survey instrument. An alpha of 0.7 is considered fair, 0.8 is good, and 0.9 is excellent. The statistical analysis was conducted with SPSS 20 Statistical Software (IBM-SPSS Version 20; SPSS Inc, Chicago, IL, USA), and  $p < 0.05$  was considered statistically significant.

Because patients were asked to simultaneously complete four different scores containing many similar questions, there was a risk of overwhelming the patients leading to missing responses. We analyzed the missing values item per item and for the total scores and found 0 to 10.5% missing values (average 4%) with the highest 10.5% only for three items in the SF-12 MCS. Because 10% missing data for a variable is considered acceptable, we performed the analyses without further evaluation or adjustment [12].

## Results

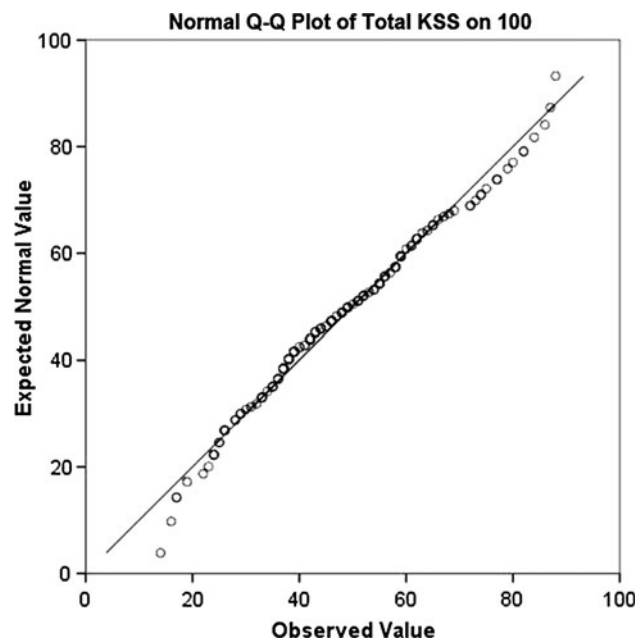
### Construct Validity

Overall, the Dutch New KSS correlated strongly with the Dutch WOMAC ( $r = -0.751$ ;  $p < 0.001$ ), Dutch KOOS ( $r = -0.723$ ;  $p < 0.001$ ), and Dutch SF-12 ( $r = 0.569$ ;  $p < 0.001$ ). Comparable domains such as pain correlated well with the KOOS Pain subscale ( $r = 0.553$ ;  $p < 0.001$ ) and the WOMAC Pain subscale ( $r = 0.556$ ;  $p < 0.001$ ). There was an even higher correlation between the Total Function domain of the Dutch New KSS and function subscales of the WOMAC ( $r = 0.759$ ;  $p < 0.001$ ) and the KOOS ( $r = 0.796$ ;  $p < 0.001$ ).

Convergent validity was additionally tested against the SF-12 PCS subscales: there was a strong and significant correlation with the SF-12 pain subscale ( $r = 0.604$ ;  $p < 0.001$ ), SF-12 physical limitations subscale ( $r = 0.655$ ;  $p < 0.001$ ), and the SF-12 physical functioning subscale ( $r = 0.521$ ;  $p < 0.001$ ). Divergent validity was confirmed by the weaker correlations with the SF-12 MCS emotional accomplishment subscale ( $r = -0.302$ ;  $p = 0.001$ ); downhearted subscale ( $r = -0.274$ ;  $p = 0.003$ ); health and social life subscale ( $r = -0.200$ ;  $p = 0.034$ ), and general health ( $r = 0.255$ ;  $p = 0.010$ ).

Content validity was proven by a significant difference between pre- and postoperative scores ( $p < 0.001$ ) in line with the KOOS, WOMAC, and SF-12 scores, which also differed significantly between the pre- and postoperative groups ( $p < 0.001$ ). Floor effects did not occur preoperatively, whereas ceiling effects were not noted at 6 months postoperatively (Fig. 1).

Of the 53 patients who received a second set of questionnaires, only 47 patients returned both sets of double questionnaires completed with a mean interval of 8.2 days (range, 2–20 days) (Table 1). The reliability proved excellent with an ICC between 0.73 and 0.92 depending on the domain tested (Table 2), which proves an adequate reproducibility. Internal consistency as indicated by Cronbach's alpha ranging from 0.84 to 0.96 for the



**Fig. 1** Distribution of the Dutch New KSS. The graph shows the normal distribution (normal Q-Q plot) of our patients' total scores using the Dutch New KSS. Q = quantile. The calculated scores were normally distributed, and no ceiling or floor effect is observed at preoperative and short-term postoperative intervals.

**Table 2.** Test-retest reliability (intraclass correlation coefficient [ICC]) and internal consistency (Cronbach's alpha [CA]) for the total score of the Dutch New Knee Society Scoring System and for the different domains

Domain	Test 1		Test 2		ICC	95% CI	CA
	Mean	SD	Mean	SD			
Symptoms (3 items)/25	12.0	5.2	11.6	5.0	0.92	0.86–0.95	0.96
Satisfaction score (5 items)/40	14.5	9.1	14.9	8.5	0.73	0.60–0.85	0.84
Expectation score (3 items)/15	11.9	2.5	11.7	2.6	0.84	0.77–0.92	0.91
Functional activity (4 items)/30	11.5	9.0	12.2	8.3	0.92	0.87–0.95	0.96
Standard activities (6 items)/30	13.3	5.7	13.9	6.1	0.83	0.76–0.91	0.91
Advanced activities (5 items)/25	4.8	5.1	4.9	4.8	0.78	0.69–0.89	0.87
Discretionary activities (3 items)/15	5.2	3.4	5.8	3.7	0.76	0.66–0.88	0.86
Total functional score (18 items)/100	35.9	18.7	37.3	18.3	0.87	0.78–0.93	0.93
Total score/180	71.9	24.6	74.0	24.1	0.82	0.69–0.90	0.93
Total score/100)	40.7	13.9	41.8	13.4	0.82	0.70–0.89	0.90

CI = confidence interval.

individual subscales was good to excellent for all domains (Table 2).

The difference in means between the Dutch New KSS of the first and the second questionnaires was 1.1 on a scale of 0 to 100. There was no significant difference between the first and second questionnaires (Wilcoxon signed rank  $p = 0.835$ ).

## Discussion

The aim of this study was to translate and validate the original New KSS for the Dutch-speaking TKA target population. None of the items in the original form caused any problems during the translation process because they were all clearly understood and interpreted by the two independent professional translators and by all patients. Because the different items of the questionnaire were universal to all patients, no cultural adaptation was necessary in the translation procedure [12, 13].

We acknowledge certain limitations of our study. First, we only collected questionnaires from patients undergoing TKA at the preoperative or short-term postoperative intervals. Therefore, the ability of the Dutch New KSS to adequately respond to further longer-term improvement could not be tested and ceiling effects could not be thoroughly assessed. Second, because we simultaneously applied four questionnaires containing several similar items, patients may have been overwhelmed, which led to missing responses. However, overall the percentage of missing data was low (average 4%). Third, patients were not very compliant to complete the retest questionnaires at the requested 1-week interval. The second Dutch New KSS scores were returned after a mean of 8.2 days (range, 2–20 days) but because a randomized controlled trial

demonstrated no difference between repeated knee scores at 2 days or 2 weeks, we retained all sets of scores [18].

The results of this study show that the Dutch version of the new KSS is a valid, reliable, and consistent instrument for Dutch-speaking patients before and after TKA. Construct validity was demonstrated by a strong and significant correlation between the total scores and subscales of the Dutch New KSS and the Dutch versions of the WOMAC, KOOS, and SF-12 scores. Because these scales, like the New KSS, are especially designed for patients with knee OA, they were considered most suitable to test the construct validity. In the validation process of the original English New KSS, comparison to SF-12 and KOOS also demonstrated strong correlations [21].

The study also demonstrated good content validity of the Dutch New KSS. There was a significant difference between preoperative and postoperative scores and none of the patients scored a minimum or maximum score, indicating that floor or ceiling effects did not occur. However, the examined patients were patients with severe OA who were candidates for a TKA or patients at short-term followup after TKA. Theoretically, a ceiling effect could occur in patients with mild OA or at long-term followup after TKA.

The test-retest reliability of the translated New KSS form was high as illustrated by a high degree of concordance (ICC) when subjects completed the Dutch New KSS twice at an approximately 1-week interval. In addition, a good to excellent internal consistency of the different test items was demonstrated with high Cronbach's alpha coefficients. Internal consistency reflects the extent to which items measure the same characteristic and reflect the individual test's true score. The Cronbach's coefficients found in this study (0.84–0.96) were comparable with the results of the original New KSS with reported values of



Cronbach's alpha ranging between 0.68 to 0.95 [21] indicating good to excellent internal consistency of the scales [6].

Compared with the other scores evaluated in this study, the Dutch New KSS performed well. The WOMAC [3] and SF-12 are considered reference scores for the assessment of patient-relevant treatment effects in OA but are not specifically responsive to knee treatment [5, 19, 20]. The KOOS [23] was developed to respond to more demanding physical functions, smaller differences in outcome, and longer-term results of treatments of a knee injury or knee OA in younger patients but has weak validity regarding the sport and recreation subscale [5]. The New KSS is intended to overcome these flaws, but it is still too early to establish whether it will become the gold standard for TKA evaluation in younger and more active patient populations.

## Conclusion

The translated Dutch New KSS was proven to be a valid and reliable instrument for the evaluation of patients with severe OA who were candidates for a TKA or patients at short-term followup after TKA. Therefore, the Dutch version is considered equivalent to the original English version [21] and can be used as a valuable measuring tool to characterize the expectations, satisfaction, symptoms, and physical activities of Dutch-speaking patients undergoing TKA.

**Acknowledgments** We thank Mieke Decokere, Lotte Verstuyft, and Eva Verplanken for data input.

## References

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, Sprangers MA, te Velde A, Verrips E. Translation, validation, and norming of the Dutch language version of the SF-36 health survey in community and chronic disease populations. *J Clin Epidemiol.* 1998;51:1055–1068.
2. Beck CT, Gable RK. Ensuring content validity: an illustration of the process. *J Nurs Meas.* 2001;9:201–213.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–1840.
4. Bland JM, Altman DG. Statistics notes. Measurement error. *Br Med J.* 1996;312:744–746.
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. Cronbach LJ, Meehl P. Construct validity in psychological tests. *Psychol Bull.* 1955;52:281–302.
7. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br.* 1998;80:63–69.
8. Englund M, Lohmander LS. Patellofemoral osteoarthritis coexistent with tibiofemoral osteoarthritis in a meniscectomy population. *Ann Rheum Dis.* 2005;64:1721–1726.
9. Franklin PD, Li W, Ayers DC. The Chitranjan Ranawat Award: Functional outcome after total knee replacement varies with patient attributes. *Clin Orthop Relat Res.* 2008;466:2597–2604.
10. Gajewski BJ, Coffland V, Boyle D, Bott M, Price LR, Leopold J, Dunton N. Assessing content validity through correlation and relevance tools. *Methodology.* 2012;8:81–96.
11. 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.
12. Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol.* 1995;24:61–63.
13. 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.
14. Haverkamp D, Breugem SJM, Sierevelt IN, Blankevoort L, van Dijk CN. Translation and validation of the Dutch version of the Oxford 12-Item Knee Questionnaire for Knee Arthroplasty. *Acta Orthop.* 2005;76:347–352.
15. Haverkamp D, Sierevelt IN, Breugem SJ, Lohuis K, Blankevoort L, van Dijk CN. Translation and validation of the Dutch version of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2006;34:1680–1684.
16. Insall J. Current concepts review: patellar pain. *J Bone Joint Surg Am.* 1982;64:147–152.
17. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, Richmond JC, Shelborne KD. Development and validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2001;29:600–613.
18. 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.
19. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (Sf-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care.* 1994;32:40–66.
20. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247–263.
21. 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.
22. Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, Willems WJ, Heyligers IC, Voaklander DC, Kelly KD, Suarez-Almazor ME. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis.* 2004;63:36–42.

23. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (Koos)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88–96.
24. 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.
25. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–233.
26. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473–483.
27. Weiss JM, Noble PC, Conditt MA, Kohl HW, Roberts S, Cook KF, Gordon MJ, Mathis KB. What functional activities are important to patients with knee replacements? *Clin Orthop Relat Res.* 2002;404:172–188.