Clinimetric quality of the new 2011 Knee Society Score: High validity, low completion rate

Remco N. Dinjens *, Rachel Senden, Ide C. Heyligers, Bernd Grimm

Atrium Medical Centre, Department Orthopaedics & Traumatology, Henri Dunantstraat 5, 6419 PC Heerlen, The Netherlands

**A B S T R A C T**

Background: The demands of the younger and more active current total knee arthroplasty (TKA) patients are not in line with the current outcome assessments. Therefore, new questionnaires are developed or adjusted, as with the popular 1989 Knee Society Score (KSS). This study is the first to investigate the clinimetric parameters of the patient-reported outcome measurement (PROM) part of the 2011 KSS.

Methods: Four-hundred-fifteen primary Dutch TKA patients were scored using the PROM part of the 2011 KSS. The scale is subdivided into an Objective (not evaluated), Satisfaction, Expectation and Function subscales. Clinimetric quality was evaluated by response and completion rate, test–retest reliability (n = 29, intraclass correlation coefficient), internal consistency (n = 172, Cronbach’s alpha), construct validity (Pearson’s correlations with 1989 KSS (n = 75) and KOOS-PS (n = 139)) and responsiveness (n = 20, paired-samples t-test, effect sizes and floor and ceiling effects).

Results: A response rate of 96% and completion rate of 43% were found. Reliability and internal consistency proved excellent with ICCs ≥ 0.79 and Cronbach’s alpha ≥ 0.76 for all subscales. Strong correlations were found between the Function subscales of the 2011 KSS and KOOS-PS (r = −0.60 to −0.83). All subscales improved significantly after intervention, with exception of Walking & Standing and Discretionary Activities. 23% reached the maximum score postoperatively in Walking & Standing, indicating a ceiling effect.

Conclusions: The 2011 KSS is a reliable, internal consistent, construct valid and responsive questionnaire to assess the outcome of the Dutch TKA patients. Optimizations (e.g. shortening the scale, simplified design) are recommended to increase the disappointing completion rate.

Clinical relevance: The 2011 KSS is a reliable, internal consistent, construct valid and responsive questionnaire to assess the outcome of the Dutch TKA patients.

© 2014 Elsevier B.V. All rights reserved.
1. Introduction

The Knee Society Scoring System (KSS) has been developed by The Knee Society as a simple rating scale to quantify the outcome of patients before and after total knee arthroplasty (TKA) [1–3]. The KSS is a clinician-administered scale which is concise and user-friendly [1–3]. It consists of a Knee Score, which only rates the knee joint itself (e.g. pain, range of motion, stability and radiographic alignment), and a Function Score (e.g. patient’s walking distance, climbing stairs and use of walking aids). Over the years, it has become widely accepted although the reliability and validity of the scale remain a subject of discussion [3–5].

The KSS was introduced in 1989 when TKA was largely performed in patients with a sedentary lifestyle [3]. Evaluation of the knee function on the basis of the patient’s ability to walk and climb stairs only was therefore acceptable [3]. However, over the last two decades the proportion of younger, more physically active patients undergoing TKA has increased [6,7]. Subsequently, these patients live longer after TKA, have higher expectations and are more demanding concerning functional outcome (e.g. stretching exercises, gardening, kneeling) [3,8]. The KSS is limited regarding these features as it measures only simple and low demanding functional aspects (stair climbing, walking etc.) [9]. Besides, studies questioned the responsiveness and reliability of the scale, which may mask functional changes over time or after intervention [4,5,10–12]. When assessing the functional outcome, the inclusion of the patient’s opinion on, for example, expectation, satisfaction and an extended set of daily activities (e.g. household, gardening, sports, playing with grandchildren) is important to evaluate the success of medical treatment [3,13].

To deal with the new generation of patients and their rising demands, current clinical scales have been optimized and new scales

---

**Satisfaction subscale (40 points)**

*How satisfied are you with the pain level of your knee while...*

1. sitting? (8 points = very satisfied)
2. lying in bed? (8 points)
3. getting out of bed? (8 points)
4. performing light household duties? (8 points)
5. performing leisure recreational activities? (8 points)

**Expectation subscale (15 points)**

*My expectations for...*

1. pain relief were... (5 points = too high)
2. being able to do my normal activities of daily living were... (5 points)
3. being able to do my leisure, recreational or sports activities were... (5 points)

**Functional Activity subscale (100 points)**

*Walking and standing (30 points)*

1. Can you walk without any aids (such as a cane, crutches etc)? (yes/no)
2. If no, which of the following aid(s) do you use? (-10 = wheelchair, -2 = brace)
3. Do you use these aid(s) because of your knees? (yes/no)
4. How long can you stand... (15 points = >1 hour)
5. How long can you walk... (15 points = >1 hour)

*...(with or without aid) before stopping as a result of knee discomfort?*

*Standard activities (30 points)*

1. Walking on a uneven underground (5 points = no bother)
2. Turning or pivoting on your leg (5 points)
3. Climbing up or down a flight of stairs (5 points)
4. Getting up from a low couch or a chair without arms (5 points)
5. Getting into or out of a car (5 points)
6. Moving laterally (stepping to the side) (5 points)

*Advanced activities (25 points)*

1. Climbing a ladder or step stool (5 points)
2. Carrying a shopping bag for a block (5 points)
3. Squatting (5 points)
4. Kneeling (5 points)
5. Running (5 points)

*Discretionary activities (15 points)*

Please check 3 of the activities below that you consider most important to you.

Nine recreational activities [swimming, gardening, etc] and 8 workout and gym activities [weightlifting, stretching exercises, etc].

*How much does your knee bother you during each of these activities?*

1. Activity A (5 points)
2. Activity B (5 points)
3. Activity C (5 points)

---

**Fig. 1. Questions for the patients in the 2011 KSS.**
have been developed. This resulted in a shift from clinician administered measures (CAMs) to patient reported outcome scales (PROMs), such as the Knee Injury and Osteoarthritis Outcome Score (KOOS), Oxford Knee Score (OKS) and Western Ontario McMaster University Osteoarthritis Index (WOMAC) [13,14]. The KSS has been revised by the Knee Society by combining patient reported questions with the currently used clinician administered questions. In addition, more demanding activities are assessed (e.g. kneeling) and the patient’s satisfaction and expectation is incorporated [2,3].

To be used as a clinical outcome tool for TKA patients, the new KSS needs to be evaluated in a large group of TKA patients. In addition, its clinimetric quality needs to be examined, which includes the practice of assessing or describing symptoms and signs by means of scales, indices and other quantitative instruments. This is the first study in which the clinimetric quality of the patient-reported outcome measures (PROM) of the 2011 KSS is investigated in a large group of TKA patients.

2. Methods

2.1. Patients

A total of 415 primary TKA patients (avg. 69 ± 9 years, 257♀: 158♂, 80 pre and 335 postoperative (follow-up 6 weeks to >5 years)) were approached. Patients indicated for TKA were addressed preoperatively by telephone or in person at the outpatient clinic during consultation and were contacted a second time, six to 12 weeks after surgery. In addition, all postoperative patients visiting the outpatient clinic for regular check-up after TKA, independent of their follow-up, were approached by mail. Every patient was asked to complete the PROM-part of the 2011 KSS. The PROM-part is the most innovative change concerning the 2011 KSS in contrast to the 1989 KSS and is therefore the main focus of investigation in this study. The objective subscale of the 2011 KSS is not evaluated in this study because it is almost identical to the 1989 KSS and comparative validity parameters for comparison are not available for the 1989 KSS. Patients who underwent a hemi-knee prosthesis or revision surgery (hemi-knee to TKA or revision of TKA) or were unable to read the Dutch language were excluded.

2.2. Questionnaires

The Dutch versions of the 2011 KSS and the Knee injury and Osteoarthritis Outcome Score-Physical Function Short Form (LK 1.0 KOOS-PS) were mailed to patients or handed over personally during their visit at the outpatient clinic [1–3,16,17]. The questionnaires were returned by mail in a pre-paid envelop or brought back by the patient at the outpatient clinic. Clinicians completed the 1989 KSS when patients visited the outpatient clinic.

2.2.1. 2011 Knee Society Score

The 2011 KSS consists of a total of 34 questions divided into four subscales which are rated separately [2,3]. (1) The clinician reported objective subscale (seven items; 100 points, not further evaluated in this study) and the patient reported subscales (2) satisfaction (five items; 40 points), (3) expectation (three items; 15 points) and (4) functional activity subscale (19 items; 100 points) [2,3]. The Functional Activity subscale is further divided into four subscales: (1) Walking & Standing (five items; 30 points), (2) standard activities (six items; 30 points), (3) advanced activities (five items; 25 points) and (4) discretionary activities (three items; 15 points) [Fig. 1]. [2,3].

The higher the score, the better the outcome in all subscales [Fig. 1]. A pre- and postoperative version of the 2011 KSS consists, showing differences in the formulation of the subscale Expectation [2,3]. In the function subscale Discretionary Activities patients have to assess three activities (e.g. running, stretching exercises, gardening etc.) that
they consider most important. Completing the questionnaire takes 10 to 30 min.

For this study the original English version was translated into Dutch using the equivalent terminology from the English and Dutch versions of the 1989 KSS and KOOS questionnaires plus a few unambiguous translations (e.g. “fietsen” for “cycling”). The 2011 KSS explicitly allows cultural adaptation of the activities, in particular the discretional activities and this was investigated for the Dutch version. However, any activity deemed culturally specific to the Dutch TKA population (e.g. ice skating) was considered less relevant than any activity in the 2011 KSS activity set including those with lowest frequency of use in the study population (e.g. golfing 1%). Thus it was possible to maintain the same list of activities which enhances the comparability between both language versions.

2.2.2. 1989 Knee Society Score (KSS)
The 1989 KSS is clinician-administered consisting of a Knee and Function Score [1]. The Knee Score rates the knee joint itself and allocates a maximum of 100 points for the evaluation of range of motion (1 point per 5°, maximum 125°), stability (medial/lateral (15 points) and anterior/posterior (10 points)) and pain (50 points) with deductions for extension lag, flexion contracture and malalignment (if leg axis <5 or >10° on radiological examination) [1]. A maximum score of 100 points represents a well-aligned knee with 125° of motion, almost none anteroposterior or mediolateral instability and no pain [1]. The Function Score considers walking distance (50 points) and stair climbing (50 points) with deduction for the use of a walking aid [1]. A patient who can walk unlimited and has no trouble with climbing stairs will obtain the maximum Function Score of 100 points [1].

2.2.3. Knee injury and Osteoarthritis Outcome Score—Physical Function Short Form (KOOS-PS)
The KOOS-PS consists of seven questions concerning physical function and is derived from the long-form KOOS using Rasch analysis [17]. It is intended to assess people’s opinions about inconveniences they experience during daily activities due to problems with their knee and has proven to be cross culturally valid [15–21]. All 7-items (e.g. rising from bed, putting on socks/stockings, rising from sitting, bending to floor, twisting/pivoting on your injured knee, kneeling and squatting) are rated from 0 to 4 points (0 represents no complaints) [17]. The rating scale is scored by summing the raw response (range 0–28) and then using a nomogram were the raw score is converted to a true interval score from 0 (no difficulty) to 100 (extreme difficulty) [17].

2.3. Statistical analysis
2.3.1. Response and completion rate
The response rate was tested by the percentage of patients who returned the questionnaire to the outpatient clinic. The percentage of fully completed questionnaires was evaluated under the returned rating scales and defined as the completion rate. For the analysis, only completely filled in questionnaires are included. Incomplete scores were examined for possible reasons which could have caused the incompleteness.

2.3.2. Internal consistency
To investigate the homogeneity of the items in the (sub)scale, internal consistency was measured using Cronbach’s alpha, with alpha > 0.70 indicating good homogeneity [22–24].
Construct validity was evaluated in a subgroup of TKA patients to investigate the correlations between the Functional Activity score of the 2011 KSS score and a) the 1989 Function score (n = 75, mean 66 (SD 9) years, 32.2%:43%, 36 pre and 39 postoperative) and b) the KOOS-PS (n = 139, mean 67 (SD 9) years, 62.2%:77%, 42 pre and 97 postoperative) were examined using Pearson’s correlations [22–24]. Evaluation of the construct validity was done by calculating Pearson’s correlations, with $r \geq 0.70$ indicating a strong correlation and $r < 0.50$ a low correlation [22].

More relations between (sub)scales were investigated like the correlation between patient satisfaction and function as measured by the 2011 KSS, 1989 KSS and KOOS.

### 2.3.5. Responsiveness

Responsiveness was evaluated by investigating the sensitivity of the 2011 KSS score to changes over time. The 2011 KSS was completed preoperatively and 6 to 12 weeks postoperative in a subgroup of 20 TKA patients (mean 64 (SD 10) years, 6/14:77) [22–24]. Responsiveness was evaluated using the paired-samples t-test and by measuring the effect size (mean score change divided by the standard deviation of the pre-operative score, score > 0.8 is considered large effect) [25,26]. For comparative reasons, the responsiveness of the KOOS-PS was examined in the same population subgroup using the same statistics. In addition, floor and ceiling effects were evaluated for all subscales of the 2011 KSS, for both pre and postoperative scales. Floor and ceiling effects refer to specific limitations encountered in questionnaires when measuring clinical outcome. A floor effect means that a patient scores at or near the minimum possible score. Ceiling effects occur when the maximum possible score is reached and no further improvement can occur. Changes in health status are an important outcome and floor and ceiling effects can influence the results. It can affect the ability of the questionnaire to detect changes over time. A ceiling and floor effect is present if $>15\%$ of the patients achieved respectively the highest or lowest possible score [23].

Data were analyzed using SPSS statistical software (version 19.0). The level of significance was set at a p-value $\leq 0.05$ for all statistical procedures.

### 3. Results

#### 3.1. Response and completion rate

The 2011 KSS produced a response rate of 96% and a completion rate of 43%. This resulted in 172 complete scores, which were used for further analysis. The low completion rate was mainly due to missing answers in the function subscales Advanced and Discretionary Activities (Fig. 2). In addition, the fact that the options “not applicable” or “I never do this” were either not available (satisfaction) or not related to a value for scoring (activities) contributed to the low completion rate.

#### 3.2. Internal consistency

A Cronbach’s alpha $\geq 0.76$ was found for all subscales in the 2011 KSS, indicating good internal consistency of all items in the subscales (Table 1).
3.3. Reliability

The mean scores for the difference between the retest and baseline values in all subscales are almost identical and every subscale showed high reliability (ICCs ≥ 0.79) as is seen in Table 2. No difference between the both repeated measurements was observed in the Wilcoxon signed rank test, further indicating good reliability (Table 2).

3.4. Construct validity

High correlations (r-range −0.74 to −0.83) were found between the KOOS-PS and the function subscales Standard and Advanced Activities indicating good construct validity (Fig. 3, Table 3). Moderate correlations (r-range −0.60 to −0.69) were found between the KOOS-PS and the function subscales Walking & Standing and Discretionary Activities (Table 3).

Moderate to low correlations (r-range 0.40 to 0.59) were found between the 1989 KSS Function Score and the function subscales of the 2011 KSS (Fig. 4, Table 3).

3.5. Responsiveness

Most subscales showed significant improvements (range 17%–32%) after TKA, except the function subscales Standing & Walking and Discretionary Activities (Fig. 6). Similar improvements were found in the KOOS-PS (range 24%, effect size 1.26, Fig. 5, Table 4). The most responsive subscale was the Satisfaction subscale showing an effect size of 2.17; in contrast to the subscale Walking & Standing which proved least responsive (effect size 0.57, Table 4).

No floor or ceiling effects were found preoperatively (Table 5). Postoperatively, a ceiling effect (23%) was found for the function subscale Walking & Standing, no floor effects were encountered (Table 5). A remarkable finding was that 36% of all pre-operative patients reached the highest possible score in the subscale Expectation, indicating that these patients expected maximum benefits from TKA.

4. Discussion

The current study investigated the clinimetric quality of the recently developed 2011 KSS in a large group of Dutch TKA patients. The 2011 KSS proved to be a reliable, internal consistent, construct valid and responsive rating scale for TKA patients at various time points of clinical follow-up. In contrast to most of the currently used rating scales, the 2011 KSS incorporates patient satisfaction and expectation and includes more demanding activities for measuring the outcome after TKA. In addition, the 2011 KSS personalizes the outcome scale for each patient by giving patients the opportunity to fill in their own three most important activities in daily life (Discretionary activities). In this way, the 2011 KSS gives additional information concerning the outcome after TKA which are in line with the demands of the younger and more demanding TKA patients. While cultural adaptation of the discretionary activity set was deemed unnecessary or even undesirable for the Dutch population, such adjustments may be an important and very useful feature of the 2011 KSS for populations with larger cultural differences to Western countries such as Asia, Middle East or Africa, defining an advantage of the 2011 KSS over other, culturally non-flexible PROMs.
The reliability of the 2011 KSS rated “good” as was shown by examination of the mean scores of the difference between retest and baseline values and a ICC ≥ 0.79 in all subscales. This is comparable to the reliability figures reported for the Dutch versions of the KOOS (ICC ≥ 0.74) but less than the International Knee Documentation Committee Rating System (IKDC) Subjective Knee Form (ICC 0.96) and Oxford 12-item knee questionnaire (ICC 0.97) [27–29]. The good reliability indicates that the 2011 KSS is consistent across repeated measurements if no changes in the patient’s perception (e.g. knee status) have occurred. It also implies that patients will interpret the questions of the rating scale every time in the same manner.

An excellent internal consistency was found for all subscales (Cronbach’s alpha ≥ 0.90), only Walking & Standing and Discretionary Activities gave slightly lower values (Cronbach’s alpha 0.76 and 0.85). This indicates good homogeneity between the items of a subscale and that the results of these items are related to each other. Comparable internal consistency was reported for the KOOS (Cronbach’s alpha ≥ 0.71), IKDC Subjective Knee Form (Cronbach’s alpha 0.92) and Oxford 12-item knee questionnaire (Cronbach’s alpha 0.94) [27–30].

The Function scale of the 2011 KSS showed good construct validity, as it measures a similar construct as other validated functional outcome scale like the KOOS-PS and the 1989 KSS Function scale (r-range 0.40–0.83). This is in line with the construct validity published for the IKDC Subjective Knee Form (r-range −0.62 to −0.77) and Oxford 12-item knee questionnaire (r-range 0.48–0.81) [28–30].

The 2011 KSS also proved to be responsive for the subscales Satisfaction and Function-Standard & Advanced Activities. No significant improvement in Function subscales Walking & Standing and Discretionary Activities were found postoperatively. One reason could be that Discretionary Activities need more time (>12 weeks) to improve after surgery because it contains more demanding activities. In many cases patients could still walk for a substantial distance preoperatively which could be a possible explanation for the lack of responsiveness in this subscale.

Despite the aim to create a new score for a more demanding population, a ceiling effect was measured for the 2011 KSS in the Function subscale Walking & Standing at postoperative assessment. At 6–12 weeks after TKA 23% of the patients had reached the maximum possible score and therefore cannot further improve any further in this subscale over time. However, ceiling effects are also reported for the KOOS [25,27] (e.g. ceiling effect in the subscale pain (range 15–28%)). Although the ceiling effect in the Function subscale Walking & Standing, this task stays an important outcome parameter because it’s the most common way of locomotion for people and underperformance in this category may indicate a red flag in the follow-up control.

Despite a very high response rate (96%), an unsatisfactory completion rate (43%) was found for the 2011 KSS. The response rate of the 2011 KSS is higher than reported for KOOS and WOMAC which showed a response rate of respectively 68% and 77% [31,32]. A possible explanation for this difference could be that in this study patients were not only contacted by mail, as in these other studies, but also personally contacted at the outpatient clinic, showing the high relevance of good doctor–patient communication when trying to achieve high compliancy with mailed PROM questionnaires.

In contrast, the completion rates reported for the KOOS and WOMAC, 95% and 94% respectively, are much higher than the completion rate of the 2011 KSS [31,32]. For the KOOS-PS, a completion rate of 75% was found in this study. One reason for this difference may be the alternative approach to deal with missing answers. Paradowski et al. excluded scores only if more than two items were missing in each of the five KOOS subscales and only one subscale had to be complete to be counted towards the completion rate [32]. This could be a possible explanation for the difference in completion rate between the new 2011 KSS and the KOOS, because this study only examined fully completed scores of the 2011 KSS. However, in comparison with the completion rate of the WOMAC (e.g. links et al. only examined the completion rate if all items were completed), the 2011 KSS remains inferior [31]. This may be due to the fact that the 2011 KSS is too time consuming (34 items vs. 24 items of the WOMAC) or too confusing in design or layout. The WOMAC and KOOS score have a more consistent layout of the questions and answers which may aid elderly patients. This particularly applies to the Function subscale Discretionary Activities with a suboptimal design logic and thus in fact being the most unanswered questions (32%) (Fig. 1). Incomplete scores (in 21% and 38% of all patients) were also due to the answer option “I never do this” (21% of patients in subscale standard activities, 38% of patients in subscale advanced activities) and as this option is not associated to a value for scoring no final score can be calculated (Fig. 1). These weaknesses could easily be resolved by a graphical redesign and firm rules on scoring the answer option “I never do this”.

To improve the completion rate, several solutions were investigated. The answer option “I never do this” was associated with the worst possible score (zero). Answer options which were left open in the subscales of the 2011 KSS were calculated following the “1998 rule” for missing answers advised for the KOOS (www.koos.nu). If two or less questions are left open in a subscale (or one question in a three item subscale), a mean average score can be calculated from the remaining completed answers in the same specific subscale. Using these rules, the completion rate improved to 72%, while the internal consistency (ICC ≥ 0.77), construct validity (r-range −0.56 to −0.76 with KOOS-PS) and responsiveness (significant improvement in the subscale satisfaction and function subscale standard activities (effect sizes 2.20 and 1.60)) remained similar. Although the completion rate is improved (from 43% to 72%), the rating scale remains inferior to the completion rates of the KOOS and WOMAC and therefore additional improvements, such as shortening the rating scale, or simplifying the design may further increase the completion rate.

This study is not without limitations. Reliability and validity testing was based on a small sample size. This was partially due to the low completion rate of the questionnaire. However, because of the small sample size, the results should be interpreted with caution. Future studies including more TKA patients are required.

### Table 4

<table>
<thead>
<tr>
<th>2011 KSS subscales</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>2.17</td>
</tr>
<tr>
<td>Function: walking &amp; standing</td>
<td>0.57</td>
</tr>
<tr>
<td>Function: standard activities</td>
<td>1.78</td>
</tr>
<tr>
<td>Function: advanced activities</td>
<td>1.04</td>
</tr>
<tr>
<td>Function: discretionary activities</td>
<td>0.78</td>
</tr>
<tr>
<td>Function total</td>
<td>1.22</td>
</tr>
<tr>
<td>KOOS-PS</td>
<td>1.26</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Floor and ceiling effects of the subscales of the 2011 KSS: percentage of patients displaying worst possible (floor effect)/best possible score (ceiling effect).</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011 KSS subscales</td>
</tr>
<tr>
<td>Satisfaction</td>
</tr>
<tr>
<td>Expectation</td>
</tr>
<tr>
<td>Function: walking &amp; standing</td>
</tr>
<tr>
<td>Function: standard activities</td>
</tr>
<tr>
<td>Function: advanced activities</td>
</tr>
<tr>
<td>Function: discretionary activities</td>
</tr>
<tr>
<td>Function total</td>
</tr>
</tbody>
</table>

---

5. Conclusion

The 2011 KSS is a reliable, internal consistent, construct valid and responsive rating scale for the younger and more demanding Dutch TKA patients. However, the 2011 KSS in its current form gave an unsatisfactory
completion rate. This can be partially improved by associating the “not applicable” option with the worst score and by interpolating a missing value when answers are left open in a subscale. Shortening the questionnaire and redesigning the layout may further raise the completion rate. Future studies evaluating the New KSS for outcome assessment shall investigate the tool’s responsiveness in longitudinal studies and its correlation with objective functional tests such as performance tests (e.g. Get-up and Go test) or motion analysis (e.g. Accelerometer based gait-test) and with patient activity, e.g. using ambulant sensors for monitoring [9,33].

References