

# Translation and Psychometric Analysis of the Indonesian Versions of the Lysholm and Tegner Scores for Patients With Anterior Cruciate Ligament Injuries

Romy Deviandri,<sup>\*†‡§</sup> MD, Hugo C. van der Veen,<sup>†</sup> MD, PhD, Andri M.T. Lubis,<sup>||</sup> MD, PhD, Maarten J. Postma,<sup>¶#\*\*††</sup> PhD, and Inge van den Akker-Scheek,<sup>†</sup> PhD

*Investigation performed at the Department of Orthopedics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands*

**Background:** The Lysholm knee score and Tegner activity scale are frequently used patient-reported outcome measures in patients with anterior cruciate ligament (ACL) injuries because of their excellent psychometric properties. These questionnaires were originally developed in the English language.

**Purpose:** To translate and cross-culturally adapt these measures into the Indonesian language and study their validity and reliability so that they can be used in the Indonesian-speaking population with ACL injuries.

**Study Design:** Cohort study (diagnosis/symptom prevalence); Level of evidence, 2.

**Methods:** After a forward-backward translation procedure and cross-cultural adaptation, validity and reliability were investigated. A total of 253 patients with an ACL injury were sent 4 questionnaires (36-Item Short Form Health Survey, Kujala anterior knee pain scale, Indonesian Lysholm knee score [I-LK], and Indonesian Tegner activity scale [I-TS]). The responses of those patients were analyzed. Following COSMIN guidelines, construct validity, test-retest reliability, internal consistency, floor and ceiling effects, and measurement errors were determined. The Bland-Altman method was used to explore absolute agreement.

**Results:** A total of 106 patients (42% response rate) were included in this study. Construct validity was considered to be good, as more than 75% of the predefined hypotheses on correlations between the I-LK, I-TS, and other measures were confirmed. Reliability proved excellent, with a high test-retest correlation for both questionnaires (intraclass correlation coefficient = 0.99). Bland-Altman analysis showed no systematic bias between testing and retesting. The internal consistency of the I-LK was good (Cronbach alpha = 0.73). For the I-LK and I-TS, floor and ceiling effects were less than 15% (floor: 0% and 4.7%, respectively; ceiling: 12.3% and 3.8%, respectively); the standard error of measurement was 1.8 and 0.9, respectively; the minimal detectable change at the individual level was 5.1 and 0.6, respectively; and the minimal detectable change at the group level was 2.4 and 0.5, respectively.

**Conclusion:** Both the I-LK and I-TS appear to be good evaluation tools for Indonesian-speaking patients with an ACL injury.

**Keywords:** ACL tear; PROMs; reliability; validity; Indonesian translation

In most countries, the annual incidence of anterior cruciate ligament (ACL) injuries is between 0.03% and 0.04%.<sup>17</sup> An ACL injury can be managed by either ACL reconstruction (ACLR), ACL repair, or nonoperative treatment.<sup>21,22</sup> The goal of these treatment options is to regain good functional stability of the knee so that patients can return to daily activities, work, and even sports.<sup>15,22</sup> To achieve this goal, the outcome of treatment should be monitored regularly. Patient-reported outcome measures (PROMs) are used to

follow up on the patient's recovery from the perspective of the patient.<sup>28</sup>

The Lysholm knee score and Tegner activity scale are frequently used PROMs to assess ACL treatment outcomes.<sup>4</sup> The Lysholm knee score was developed to determine the functional status of the patient after ACL treatment.<sup>24</sup> The Tegner activity scale was developed as an extension of the Lysholm score and designed to assess activity levels based on work and sports activities.<sup>4</sup> Both have been translated and culturally adapted into many languages in different cultural settings,<sup>9,18,29</sup> but this has not yet been done for Indonesian-speaking patients, with a total population of 270 million.<sup>6</sup>

The Orthopaedic Journal of Sports Medicine, 10(1), 23259671211066506  
DOI: 10.1177/23259671211066506  
© The Author(s) 2022

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

Hence, the purpose of this article was to translate and cross-culturally adapt the English versions of the Lysholm knee score and Tegner activity scale into Indonesian versions as well as to study the validity and reliability of these questionnaires in patients with an ACL injury. We hypothesized that the Indonesian version of these PROMs would be both valid and reliable for use in the Indonesian-speaking population with ACL injuries.

## METHODS

Patients gave informed consent before participating in this study. The study was approved by a local institutional review board (B/008/UN 19.5.1.1.8/KEPKK/2020).

### Translation Procedure

The English versions of the Lysholm knee score and Tegner activity scale were translated into Indonesian using the procedure of forward-backward translation as originally proposed by Guillemin et al<sup>10</sup> and revised by Beaton et al.<sup>2</sup> This procedure comprises the following steps: initial translation, translation synthesis, back translation, committee review, pretesting, and submission and appraisal of all written reports to the committee. The expert committee, consisting of 3 sports medicine orthopaedic surgeons (A.M.T.L., R.D., and M.P.), 1 methodological expert, and 1 translator, drafted the final Indonesian versions of the Lysholm knee score and Tegner activity scale, which were both pretested on 10 patients with an ACL injury. One of the researchers (R.D.) documented any difficulties that patients experienced while completing the questionnaires and then processed the questionnaires into their final versions.

### Patients and Procedures

A total of 253 randomly selected patients were invited by email to take part in this study between April and July 2020. All were Indonesian patients treated for an ACL injury with ACLR between January 2015 and March 2020 at a hospital in Indonesia. We selected the patients from the data of the ACL member community, a nonprofit organization consisting of patients treated with ACLR in Indonesia. We used the systematic random sampling method by selecting a patient with odd registry numbers as samples. After

giving informed consent, participants received 2 sets of questionnaires (parts A and B) by email. Part A was sent in the first week and included the Indonesian versions of the Lysholm knee score (I-LK), Tegner activity scale (I-TS), 36-Item Short Form Health Survey (SF-36), and Kujala anterior knee pain scale. Participants were asked to complete and immediately return the I-LK and I-TS again after 1 week to assess test-retest reliability. To determine whether health status and knee function remained stable between completion of both parts A and B, a global rating of change question to determine a change of status since filling out the previous questionnaire was provided. The 3 possible responses were (1) no; (2) yes, the problem shifted for the better; and (3) yes, the problem shifted for the worse. Participants who reported being much better or getting worse were excluded from test-retest reliability analysis; patients who returned both part A and part B questionnaires on the same day or more than 1 month apart were excluded as well.

### Patient-Reported Outcome Measures

The Lysholm knee score is a region-specific PROM. It includes 8 questions and is designed to measure perceived symptoms and function in patients with a variety of knee conditions. Questions concern limping, locking, pain, stair climbing, use of support, instability, swelling, and squatting. The questionnaire is scored by summing the scores for the individual items and then transforming the score to a scale that ranges from 0 to 100, with 100 indicating an absence of symptoms and higher levels of functioning.<sup>24</sup>

The Tegner activity scale is designed to assess activity levels based on work and sports activities. This scale consists of a numerical range from 0 to 10, and each value indicates the ability to perform specific activities, with the highest value indicating the best score. The Tegner activity scale was developed as an addition to the Lysholm knee score.<sup>24</sup>

The Kujala anterior knee pain scale is a disease-specific outcome measure for patellofemoral disorders of the knee. It consists of 13 questions, with a total score ranging from 0 to 100 and the highest value indicating the best score. Questions are asked regarding whether there is pain while walking, squatting, running, or jumping or during prolonged sitting with the knee in flexion; whether there is limping, swelling, or subluxation of the patella; whether there is a need for a walking aid; and whether there is

\*Address correspondence to Romy Deviandri, MD, Faculty of Medicine, Universitas Riau, Diponegoro Street No. 1, Pekanbaru, Indonesia 28133 (email: r.deviandri@rug.nl; romydeviandri@lecturer.unri.ac.id).

†Department of Orthopedics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

‡Department of Physiology, Faculty of Medicine, Universitas Riau, Pekanbaru, Indonesia.

§Division of Orthopedics-Sports Injury, Arifin Achmad Hospital, Pekanbaru, Indonesia.

||Department of Orthopedics, Faculty of Medicine, Universitas Indonesia/Cipto Mangunkusumo Hospital, Jakarta, Indonesia.

¶Department of Health Sciences, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

#Department of Economics, Econometrics & Finance, University of Groningen, Faculty of Economics & Business, Groningen, the Netherlands.

\*\*Department of Pharmacology and Therapy, Universitas Airlangga, Surabaya, Indonesia.

††Center of Excellence in Higher Education for Pharmaceutical Care Innovation, Universitas Padjadjaran, Bandung, Indonesia.

Final revision submitted August 26, 2021; accepted September 27, 2021.

The authors have declared that there are no conflicts of interest in the authorship and publication of this contribution. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

existing atrophy in the quadriceps muscle, flexion deficiency, or pain.<sup>13</sup>

The SF-36 is a generic measure used to examine the quality of life. It consists of 8 components, including physical functioning, physical role functioning, bodily pain, general health perceptions, vitality, social functioning, emotional role functioning, and mental health. Standardized scores range from 0 to 100, with higher scores indicating a better health status. In addition, the sum of the physical functioning, physical role functioning, bodily pain, and general health perceptions subscales generates a physical component summary (PCS), while the sum of the vitality, social functioning, emotional role functioning, and mental health subscales generates a mental component summary (MCS).<sup>27</sup>

### Validity

Validity is the extent to which a scale measures what it is supposed to measure—that is, whether it has the intended interpretation. The construct validity of the I-LK and I-TS is measured by determining the correlation with other questionnaires. A total of 10 predefined hypotheses were formulated about the magnitude of the correlation between the I-LK, I-TS, each subscale of the SF-36, and Kujala scale following the COSMIN guidelines.<sup>16</sup> In the Turkish version of the I-LK, Celik et al<sup>5</sup> found a correlation of 0.74 with the Kujala scale, based on which the current study expected a correlation of 0.7 or higher for the I-LK with the Kujala scale, and a correlation of 0.4 or higher for the I-TS with the Kujala scale. The I-LK and I-TS were developed to measure physical functioning rather than social and/or emotional aspects; therefore, relatively low correlations were expected between the I-LK and I-TS with the MCS of the SF-36. On the other hand, relatively high correlations were expected between the I-LK and I-TS with the PCS of the SF-36. Swanenburg et al<sup>23</sup> found correlations of 0.60 between the Lysholm score and the PCS of the SF-12, 0.02 between the Lysholm score and the MCS of the SF-12, 0.69 between the Tegner scale and the PCS of the SF-12, and 0.06 between the Tegner scale and the MCS of the SF-12. Based on the correlations found in that study, we expected a correlation of 0.4 or higher between the I-LK and the PCS of the SF-36, 0.4 or lower between the I-LK and the MCS of the SF-36, 0.4 or higher between the I-TS and the PCS of the SF-36, and 0.4 or lower between the I-TS and the MCS of the SF-36. Construct validity is considered good when at least 75% of the hypotheses are confirmed.<sup>25</sup>

### Floor and Ceiling Effects

The occurrence of floor and ceiling effects was assessed. These effects are considered to be present if more than 15% of respondents achieve the lowest or highest possible score.<sup>25</sup>

### Reliability

Reliability represents the extent to which participants can be distinguished from each other, despite measurement errors. Internal consistency, test-retest reliability, and

measurement errors were measured to determine reliability. The Bland-Altman method was used to assess absolute agreement, which reflects the amount of agreement in repeated measurements.<sup>3</sup>

### Statistical Analysis

The characteristics of the study population and scores on the questionnaires are described using means and standard deviations or frequencies and percentages. To assess construct validity, Pearson correlation coefficients—if the variable showed a normal distribution—were determined between the scores on the I-LK, I-TS, and other questionnaires. The Cronbach alpha was calculated to determine the internal consistency of the I-LK.<sup>7</sup> Values between 0.70 and 0.95 are considered to indicate good internal consistency.<sup>25</sup>

The test-retest reliability of the I-LK and I-TS was determined by measuring the intraclass correlation coefficient (ICC).<sup>16</sup> Values <0.50, 0.50 to 0.74, 0.75 to 0.90, and >0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.<sup>12</sup> The standard error of measurement (SEM) and minimal detectable change (MDC) were measured to determine measurement errors. The SEM was calculated by multiplying the pooled standard deviation by  $\sqrt{(1-r)}$ , where  $r$  is the ICC.<sup>8</sup> The formula  $1.96 \times \text{SEM} \times \sqrt{2}$  was used to measure the MDC at the individual level ( $\text{MDC}_{\text{ind}}$ ) and at the group level ( $\text{MDC}_{\text{grp}}$ ) by dividing  $\text{MDC}_{\text{ind}}$  by  $\sqrt{n}$ .<sup>25</sup>

Bland-Altman plots were used to measure absolute reliability; no systematic bias was present when 0 was in the 95% CI of the mean difference between the first and second administration of the I-LK and I-TS. The formula mean difference  $\pm 1.96 \times \text{SD}_{\text{diff}}$ , where  $\text{SD}_{\text{diff}}$  is the standard deviation of the mean difference between the first and second administration of the I-LK and I-TS, was used to determine the 95% limits of agreement.<sup>3</sup> Statistical analyses were performed using SPSS Statistics Version 26.0 (IBM), with a level of significance set at 5%.

## RESULTS

### Translation and Cross-Cultural Adaptation

There were some cultural adaptations during the procedure of translation. The word *squatting* in the Lysholm score was changed to *squat during prayer*, which is quite common among the Indonesian population. The words *bandy*, *ice hockey*, and *cross country skiing* in the Tegner scale were removed because these kinds of sports are not typical in Indonesia. The activity downhill skiing was changed to mountain biking, and we added taekwondo, volleyball, and the typical Indonesian sports *sepak takraw* (foot volleyball) and *silat* (a martial art).

### Patient Characteristics

Of the 253 patients who were invited to participate, 106 (42%) responded and returned the first mailing of

TABLE 1  
Characteristics of the Study Participants<sup>a</sup>

Characteristic	Validity and Internal Consistency Analysis Group (n = 106)	Test-Retest Reliability Analysis Group (n = 75)
Age, y	30.5 ± 9.2	30.0 ± 9.2
Sex		
Male	75 (70.8)	48 (64.0)
Female	31 (29.2)	27 (36.0)
Affected side		
Right	60 (56.6)	43 (57.3)
Left	46 (43.4)	32 (42.7)
Both	0 (0.0)	0 (0.0)
Diagnosis		
Isolated ACL injury	103 (97.2)	72 (96.0)
Multiligament injury	3 (2.8)	3 (4.0)
Occupation		
Professional athlete	14 (13.2)	13 (17.3)
Nonathlete	92 (86.8)	62 (82.7)
Activity when injury occurred		
Activity of daily living	11 (10.3)	5 (6.7)
Sport	73 (68.9)	63 (84.0)
Traffic-related accident	20 (18.9)	6 (8.0)
Work	2 (1.9)	1 (1.3)

<sup>a</sup>Data are reported as mean ± SD or n (%). ACL, anterior cruciate ligament.

questionnaires. Of these 106 patients, 75 (70.8%) completed and returned the second mailing. None of the patients were excluded because of a change in functional status when they filled out the questionnaires the second time (as measured with the global rating of change question) or because of missing data. Hence, data of 106 patients were used for validity and internal consistency analyses, and data of 75 patients were used for test-retest reliability analysis. Most patients were male nonathletes with a mean age of 30.5 ± 9.2 years. The mean scores of the first and second assessments of the I-LK were 73.4 ± 18.5 and 73.3 ± 18.5, respectively, and of the I-TS were 4.7 ± 2.7 and 4.8 ± 2.7, respectively. Patient characteristics are described in detail in Table 1. Descriptive statistics of the PROMs can be found in Table 2.

### Validity

Of the 10 predefined hypotheses on the magnitude of the correlation between the I-LK, the I-TS, and either the Kujala scale or SF-36, 8 hypotheses were confirmed (total of 80%) (Table 3). The I-LK showed a strong correlation with the Kujala scale ( $r = 0.81$ ) and with the PCS of the SF-36 ( $r = 0.66$ ). The I-TS showed a moderate correlation with the Kujala scale ( $r = 0.47$ ) and with the PCS of the SF-36 ( $r = 0.57$ ). As expected, correlation coefficients between

TABLE 2  
PROM Scores<sup>a</sup>

Measure	Mean ± SD
Indonesian Lysholm knee score	
Assessment 1	73.4 ± 18.5
Assessment 2	73.3 ± 18.5
Indonesian Tegner activity scale	
Assessment 1	4.7 ± 2.7
Assessment 2	4.8 ± 2.7
Kujala anterior knee pain scale	81.6 ± 14.9
SF-36	
Physical functioning	68.5 ± 26.6
Physical role functioning	35.8 ± 42.0
Bodily pain	72.1 ± 24.2
General health perceptions	67.5 ± 18.3
Vitality	65.1 ± 20.6
Social functioning	68.6 ± 21.1
Emotional role functioning	46.5 ± 45.5
Mental health	69.9 ± 22.2
PCS	43.6 ± 9.7
MCS	35.2 ± 16.6

<sup>a</sup>MCS, mental component summary; PCS, physical component summary; PROM, patient-reported outcome measure; SF-36, 36-Item Short Form Health Survey.

TABLE 3  
Predefined Hypothesized and Actual Correlation Between PROMs<sup>a</sup>

Hypothesis	Correlation	Hypothesized	Actual
(1) I-LK and Kujala scale		>0.7	0.81 <sup>b</sup>
(2) I-LK and SF-36 PCS		>0.4	0.66 <sup>b</sup>
(3) I-LK and SF-36 MCS		<0.4	0.62
(4) I-TS and Kujala scale		>0.4	0.47 <sup>b</sup>
(5) I-TS and SF-36 PCS		>0.4	0.57 <sup>b</sup>
(6) I-TS and SF-36 MCS		<0.4	0.33 <sup>b</sup>
(7) I-LK and Kujala scale stronger than I-LK and SF-36 PCS			0.81:0.66 <sup>b</sup>
(8) I-LK and SF-36 PCS stronger than I-LK and SF-36 MCS			0.66:0.62 <sup>b</sup>
(9) I-TS and Kujala scale stronger than I-TS and SF-36 PCS			0.47:0.57
(10) I-TS and SF-36 PCS stronger than I-TS and SF-36 MCS			0.57:0.33 <sup>b</sup>

<sup>a</sup>I-LK, Indonesian Lysholm knee score; I-TS, Indonesian Tegner activity scale; MCS, mental component summary; PCS, physical component summary; PROM, patient-reported outcome measure; SF-36, 36-Item Short Form Health Survey.

<sup>b</sup>Hypothesis was confirmed.

the I-TS and the MCS of the SF-36 were less than 0.4 ( $r = 0.33$ ). Correlation coefficients between the I-LK and the MCS of the SF-36 were higher than expected ( $r > 0.62$ ). The I-LK was more strongly related to the PCS than to the MCS of the SF-36 ( $r = 0.66$  and  $r = 0.62$ , respectively) (Table 4). The floor and ceiling effects analysis revealed that minimal and maximal possible values were less than 15% of all values given (for I-LK and I-TS: floor: 0% and 4.7%, respectively; ceiling: 12.3% and 3.8%, respectively).

Internal Consistency

The Cronbach alpha of the I-LK was 0.73, indicating good internal consistency.

Test-Retest Reliability

Both the I-LK and I-TS had an ICC value of 0.99 ( $P < .001$ ) between the test and retest scores, and the 95% CI ranged from 0.99 to 1.00.

TABLE 4  
Pearson Correlation Between PROMs<sup>a</sup>

Measure	Indonesian Lysholm Knee Score	Indonesian Tegner Activity Scale
SF-36		
Physical functioning	0.67	0.48
Physical role functioning	0.65	0.48
Bodily pain	0.62	0.50
General health perceptions	0.54	0.45
Vitality	0.51	0.36
Social functioning	0.43	0.29
Emotional role functioning	0.59	0.34
Mental health	0.55	0.33
PCS	0.66	0.57
MCS	0.62	0.33
Kujala anterior knee pain scale	0.81	0.47

<sup>a</sup>MCS, mental component summary; PCS, physical component summary; PROM, patient-reported outcome measure; SF-36, 36-Item Short Form Health Survey.

Measurement Error

The SEM, MDC<sub>ind</sub>, and MDC<sub>grp</sub> of the I-LK were calculated to be 1.8, 5.1, and 0.6, respectively. Bland-Altman analysis revealed a mean difference between the 2 I-LK assessments of 0.1, with the 95% CI ranging from 4.0 to 4.2 (95% limit of agreement = 36.1-36.2) (Figure 1). No systematic bias was present because a value of 0 was in the 95% CI of the mean difference between the test and retest scores.

The SEM, MDC<sub>ind</sub>, and MDC<sub>grp</sub> of the I-TS were calculated to be 0.9, 4.3, and 0.5, respectively. Bland-Altman analysis revealed a mean difference between the 2 I-TS assessments of 0.01, with the 95% CI ranging from 5.09 to 5.11 (95% limit of agreement = 5.38-5.40) (Figure 2). No systematic bias was presented because a value of 0 was in the 95% CI of the mean difference between the test and retest scores.

DISCUSSION

This study aimed to translate both the Lysholm knee score and the Tegner activity scale into Indonesian and gain insight into the validity and reliability of those questionnaires in an Indonesian-speaking population with an ACL injury. Based on the results, both questionnaires can be considered valid and reliable for use in Indonesian patients after an ACL injury.

The construct validity of both the I-LK and I-TS can be considered good, with more than 75% of predefined hypotheses confirmed. As hypothesized, the I-LK showed a strong correlation with the Kujala scale ( $r = 0.81$ ). This is comparable with the Turkish version ( $r = 0.78$ ).<sup>5</sup> As hypothesized,

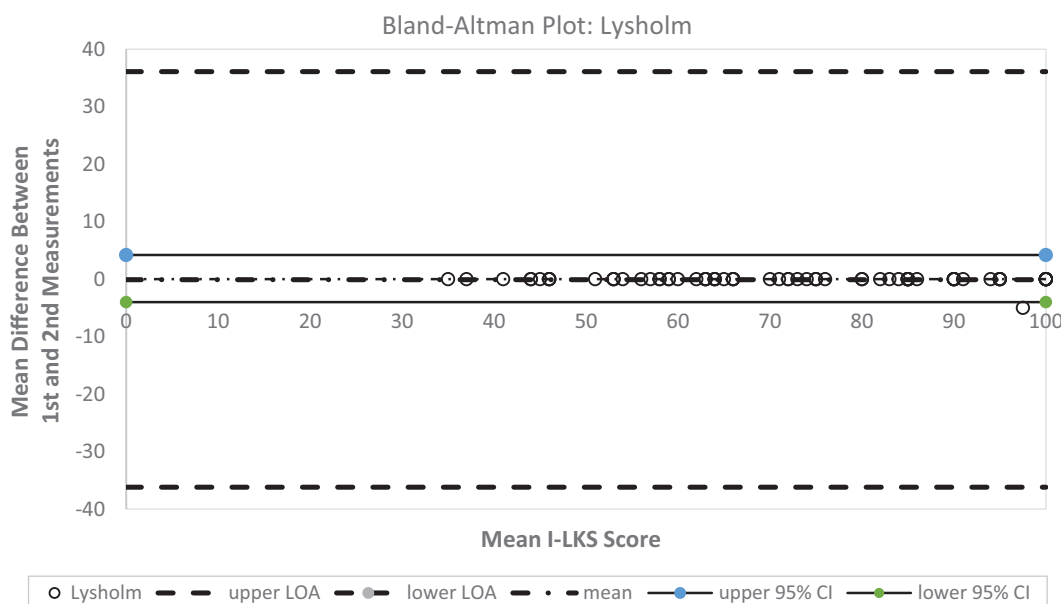
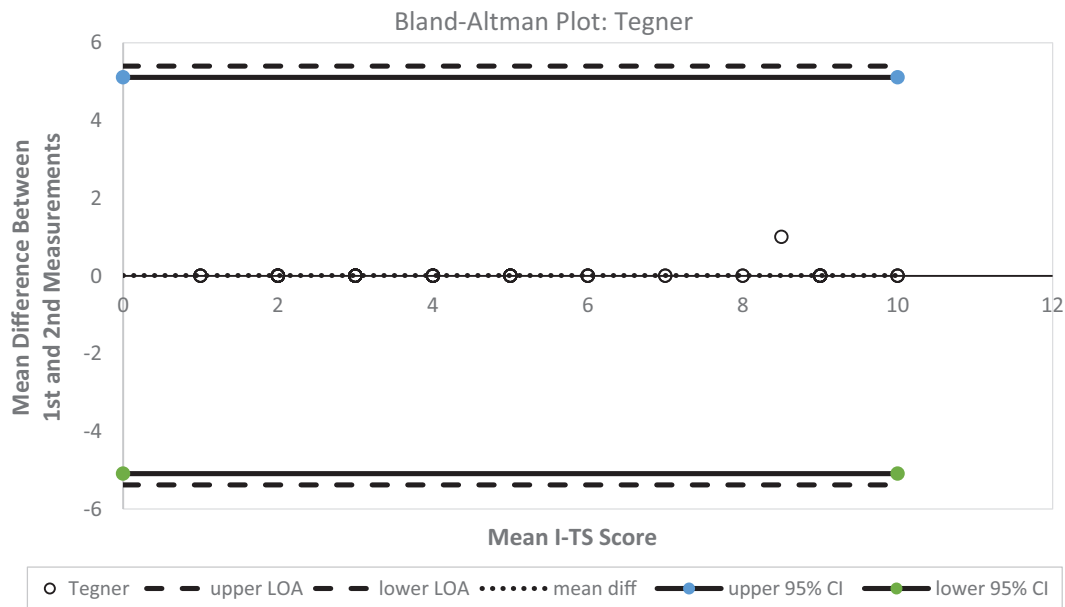


Figure 1. Bland-Altman plot of the Indonesian Lysholm knee score (I-LK), with the 2 dotted lines representing the 95% CI and the 2 dashed lines representing the limit of agreement (superior and inferior) and mean difference (intermediate).



**Figure 2.** Bland-Altman plot of the Indonesian Tegner activity scale (I-TS), with the 2 dotted lines representing the 95% CI and the 2 dashed lines representing the limit of agreement (superior and inferior) and mean difference (intermediate).

the I-LK showed a strong correlation with the PCS of the SF-36 ( $r = 0.66$ ). These results are in line with those of several other versions, such as the English ( $r = 0.68$ ),<sup>14</sup> German ( $r = 0.60$ ),<sup>23</sup> and Turkish ( $r = 0.57$ ) language versions.<sup>5</sup> Although contrary to our hypothesis that the correlation of the I-TS with the Kujala scale was lower than that with the SF-36 PCS, both correlations were higher than 0.40, as hypothesized. The second finding contradicting our hypothesis was that the I-LK showed a correlation higher than 0.40 ( $r = 0.62$ ), with the MCS of the SF-36. Still, the I-LK was more strongly correlated to the PCS than to the MCS of the SF-36, confirming our hypothesis. The correlation between the I-LK and the MCS of the SF-36 in this study was substantially higher than that found in other studies, for example, in the German ( $r = 0.02$ )<sup>18</sup> and Turkish ( $r = 0.14$ )<sup>5</sup> versions. We postulate that these results were obtained because of the characteristics of our sample—young athletes and active patients whose physical condition could have a substantial impact on their mental state.<sup>19</sup> The studies on the Turkish and German versions both included an older population of persons with various additional conditions such as knee osteoarthritis, patellofemoral pain, and meniscal lesions, which might explain the lower correlation between the I-LK and the MCS of the SF-36 compared to our study.

Other questionnaires such as the Hip and Knee Questionnaire, which was used in a Spanish Lysholm score study,<sup>1</sup> or the International Knee Documentation Committee subjective knee form, which was used in both a Dutch<sup>9</sup> and Chinese Lysholm score study,<sup>26</sup> could have been used to test construct validity. However, as no validated Indonesian translations are available for those

questionnaires, we could not use them to study the construct validity of the I-LK and I-TS.

Further affirming the occurrence of floor and ceiling effects of the I-LK and I-TS, an acceptable level of floor and ceiling effects was found of less than 15%.<sup>16</sup> This is lower than those effects reported in previous studies of the English<sup>4</sup> and Dutch<sup>9</sup> versions, both reporting more than 30%. This result indicates good content validity of the I-LK and I-TS. Also, the internal consistency of the I-LK can be considered good, with a Cronbach alpha of 0.73. This is comparable with the Chinese (0.73),<sup>26</sup> Turkish (0.68),<sup>5</sup> Dutch (0.83),<sup>9</sup> German (0.73),<sup>29</sup> and English (0.73)<sup>4</sup> versions.

Both the I-LK and I-TS showed excellent test-retest reliability (ICC = 0.99). Although the ICC between the 2 assessments of the I-LK was higher than that found in some of the studies on other language versions such as the Turkish (ICC = 0.82)<sup>17</sup> and German (ICC = 0.87)<sup>23</sup> versions, similar correlations implicating excellent reliability were found for the English (ICC = 0.97),<sup>11</sup> Chinese (ICC = 0.93),<sup>26</sup> and Dutch (ICC = 0.93)<sup>9</sup> versions. The time period between the 2 assessments of the I-LK and I-TS may have possibly caused these differences in the ICC. According to COSMIN guidelines, the test-retest time interval should be long enough to prevent recall bias and short enough to ensure that patient characteristics have not changed regarding the construct to be measured.<sup>16</sup> A time interval of 2 weeks used in the Turkish version<sup>5</sup> may have caused the lower ICC value compared to the that found in our study using a 1-week interval.

The I-LK and I-TS indicated sufficient capability for group comparisons, as only small values are needed to

detect change. The SEM of the I-LK was 1.8, the  $MDC_{ind}$  was 5.1, and the  $MDC_{grp}$  was 0.6. These values are comparable to those presented in the Dutch ( $SEM = 3.2$ ;  $MDC = 8.9$ )<sup>9</sup> and German ( $SEM = 7.84$ ;  $MDC = 21.72$ )<sup>23</sup> versions. Further, the SEM of the I-TS was 0.9, the  $MDC_{ind}$  was 4.3, and the  $MDC_{grp}$  was 0.5. These values also are comparable to those presented in the Dutch ( $SEM = 0.4$ ;  $MDC = 1.2$ )<sup>9</sup> and German ( $SEM = 0.65$ ;  $MDC = 1.80$ )<sup>23</sup> versions. As only values higher than the SEM can be distinguished from measurement errors, to detect a statistically significant change in scores on the I-LK and I-TS, the difference should be higher than the SEM. To distinguish from measurement errors and confirm that a real change occurred, the difference between 2 measurements should be greater than the  $MDC_{ind}$  value in individual patients. Given that the  $MDC_{ind}$  values are 5.1 and 4.3, respectively, it is questionable whether the I-LK and I-TS are appropriate tools for monitoring individual patients over time.

### Limitations

There are some limitations to this study. The response rate was quite low (42%), but the total number of participants can be considered sufficient: According to COSMIN guidelines, at least 100 patients are needed to examine validity and 50 patients to examine test-retest reliability.<sup>25</sup> Another potential limitation is that we compared the I-LK and I-TS with only 1 other disease-specific questionnaire (Kujala scale) and a general health-related quality of life questionnaire (SF-36). However, no other validated Indonesian versions of a PROM are available to assess knee function or patients' quality of life besides these scores.

Future research should explore the responsiveness of the I-LK and I-TS as well as the minimal important change (MIC). The MIC should be determined to explore whether a measured difference is also clinically important as perceived by the patient.<sup>25</sup> For the English versions, the reported MIC of the Lysholm knee score ranged from 3.7 to 9.2,<sup>20</sup> and the effect sizes for the Lysholm knee score and Tegner activity scale were 1.1 and 1.0, respectively, with standardized response means of 1.20 and 0.96, respectively.<sup>4</sup> These findings are not transferable to the Indonesian version, though, so the MIC, effect size, and standardized response mean of the I-LK and I-TS should be assessed in future research. Information on the responsiveness and capability of the I-LK and I-TS to detect change over time is required for these measures to be used in longitudinal research and in clinical practice to monitor Indonesian-speaking patients.

### CONCLUSION

The I-LK and I-TS can both be considered valid and reliable questionnaires for use in Indonesian patients with an ACL injury. These questionnaires enable us to measure patient-perceived symptoms, function, and outcomes of treatment in the Indonesian population with ACL injuries. More research is needed into the MIC and responsiveness.

### ACKNOWLEDGMENT

The authors extend their gratitude to all members of the Indonesian Orthopedic Society for Sport Medicine and Arthroscopy (IOSSMA) who participated in the study.

### INSTITUTIONAL REVIEW BOARD STATEMENT

Number: B/102/UN19.5.11.8/UEPKK/2020. Declared to be ethically appropriate in accordance to 7 (seven) WHO 2011 Standards, 1) Social Values, 2) Scientific Values, 3) Equitable Assessment and Benefits, 4) Risks, 5) Persuasion/Exploitation, 6) Confidentiality and Privacy, and 7) Informed Consent, referring to the 2016 CIOMS Guideline. This is as indicated by the fulfillment of the indicators of each standard.

### REFERENCES

1. Arroyo-Morales M, Martin-Alguacil J, Lozano-Lozano M, et al. The Lysholm score: cross cultural validation and evaluation of psychometric properties of the Spanish version. *PLoS One*. 2019;14(8): e0221376.
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186-3191.
3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476): 307-310.
4. Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. *Am J Sports Med*. 2009;37(5):890-897.
5. Celik D, Coşkunsu D, Kılıçoğlu Ö. Translation and cultural adaptation of the Turkish Lysholm knee scale: ease of use, validity, and reliability knee. *Clin Orthop Relat Res*. 2013;471(8):2602-2610.
6. Central Bureau of Statistics of Indonesia. Portrait of the 2020 population census towards one Indonesian population data (Potret Sensus Penduduk 2020 Menuju Satu Data Kependudukan Indonesia). Published 2021. Available at: <https://www.bps.go.id/publication.html>. Accessed February 6, 2021.
7. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302.
8. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10): 1033-1039.
9. Eshuis R, Lentjes GW, Tegner Y, Wolterbeek N, Veen MR. Dutch translation and cross-cultural adaptation of the Lysholm score and Tegner activity scale for patients with anterior cruciate ligament injuries. *J Orthop Sports Phys Ther*. 2016;46(11):976-983.
10. 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(12):1417-1432.
11. Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am*. 1998;80(8):1132-1145.
12. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2): 155-163.
13. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. *Arthroscopy*. 1993; 9(2):159-163.
14. Marx RG, Jones EC, Allen AA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am*. 2001;83(10):1459-1469.

15. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med.* 2001;29(2):213-218.
16. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737-745.
17. Moses B, Orchard J, Orchard J. Systematic review: annual incidence of ACL injury and surgery in various populations. *Res Sports Med.* 2012;20(3-4):157-179.
18. Negahban H, Mostafaei N, Sohani SM, et al. Reliability and validity of the Tegner and Marx activity rating scales in Iranian patients with anterior cruciate ligament injury. *Disabil Rehabil.* 2011;33(23-24):2305-2310.
19. Nwachukwu BU, Chang B, Voleti PB, et al. Preoperative Short Form Health Survey score is predictive of return to play and minimal clinically important difference at a minimum 2-year follow-up after anterior cruciate ligament reconstruction. *Am J Sports Med.* 2017;45(12):2784-2790.
20. Ogura T, Ackermann J, Mestriner AB, Merkely G, Gomoll AH. The minimal clinically important difference and substantial clinical benefit in the patient-reported outcome measures of patients undergoing osteochondral allograft transplantation in the knee. *Cartilage.* 2021; 12(1):42-50.
21. Smith TO, Postle K, Penny F, McNamara I, Mann CJ. Is reconstruction the best management strategy for anterior cruciate ligament rupture? A systematic review and meta-analysis comparing anterior cruciate ligament reconstruction versus non-operative treatment. *Knee.* 2014; 21(2):462-470.
22. Svantesson E, Hamrin Senorski E, Webster KE, et al. Clinical outcomes after anterior cruciate ligament injury: Panther Symposium ACL Injury Clinical Outcomes Consensus Group. *Knee Surg Sports Traumatol Arthrosc.* 2020;28(8):2415-2434.
23. Swanenburg J, Koch PP, Meier N, Wirth B. Function and activity in patients with knee arthroplasty: validity and reliability of a German version of the Lysholm score and the Tegner activity scale. *Swiss Med Wkly.* 2014;144:w13976.
24. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res.* 1985;(198):43-49.
25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.
26. Wang W, Liu L, Chang X, Jia ZY, Zhao JZ, Xu WD. Cross-cultural translation of the Lysholm knee score in Chinese and its validation in patients with anterior cruciate ligament injury. *BMC Musculoskelet Disord.* 2016;17(1):1-8.
27. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care.* 1992;30(6):473-483.
28. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect.* 2017;20(1):11-23.
29. Wirth B, Liffert F, De Bruin ED. [Development and evaluation of a German version of the Lysholm score for measuring outcome after anterior cruciate ligament injuries]. *Sportverletz Sportschaden.* 2011; 25(1):37-43.