Background. Existing lower-limb, region-specific, patient-reported outcome measures have clinimetric limitations, including limitations in psychometric characteristics (eg, lack of internal consistency, lack of responsiveness, measurement error) and the lack of reported practical and general characteristics. A new patient-reported outcome measure, the Lower Limb Functional Index (LLFI), was developed to address these limitations.

Objective. The purpose of this study was to overcome recognized deficiencies in existing lower-limb, region-specific, patient-reported outcome measures through: (1) development of a new lower-extremity outcome scale (ie, the LLFI) and (2) evaluation of the clinimetric properties of the LLFI using the Lower Extremity Functional Scale (LEFS) as a criterion measure.

Design. This was a prospective observational study.

Methods. The LLFI was developed in a 3-stage process of: (1) item generation, (2) item reduction with an expert panel, and (3) pilot field testing (n=110) for reliability, responsiveness, and sample size requirements for a larger study. The main study used a convenience sample (n=127) from 10 physical therapy clinics. Participants completed the LLFI and LEFS every 2 weeks for 6 weeks and then every 4 weeks until discharge. Data were used to assess the psychometric, practical, and general characteristics of the LLFI and the LEFS. The characteristics also were evaluated for overall performance using the Measurement of Outcome Measures and Bot clinimetric assessment scales.

Results. The LLFI and LEFS demonstrated a single-factor structure, comparable reliability (intraclass correlation coefficient [2,1]=.97), scale width, and high criterion validity (Pearson r=.88, with 95% confidence interval [CI]). Clinimetric performance was higher for the LLFI compared with the LEFS on the Measurement of Outcome Measures scale (96% and 95%, respectively) and the Bot scale (100% and 83%, respectively). The LLFI, compared with the LEFS, had improved responsiveness (standardized response mean=1.75 and 1.64, respectively), minimal detectable change with 90% CI (6.6% and 8.1%, respectively), and internal consistency (α=.91 and .95, respectively), as well as readability with reduced user error and completion and scoring times.

Limitations. Limitations of the study were that only participants recruited from outpatient physical therapy clinics were included and that no specific conditions or diagnostic subgroups were investigated.

Conclusion. The LLFI demonstrated sound clinimetric properties. There was lower response error, efficient completion and scoring, and improved responsiveness and overall performance compared with the LEFS. The LLFI is suitable for assessment of lower-limb function.
P

atent-reported outcome measures of function are critical for assessing musculoskeletal conditions. Function is the level of activities an individual performs to realize the needs of daily living. Numerous lower-limb patient-reported outcome measures assess function for specific joints, joint conditions, or region-specific conditions. However, there is limited consensus regarding which tools to use. Consequently, a need exists for a simple, reliable, and valid tool that effectively measures patient-rated lower-limb function.

Five tools were designed to measure the lower limb as a single regional kinetic chain. Only the Lower Extremity Functional Scale (LEFS) and the Foot and Ankle Ability Measure (FAAM) have detailed their clinimetric properties, but both have limitations and neither have demonstrated readability, completion time, and scoring time. Clinimetric properties include psychometric, practical, and general characteristics. (eAppendix 1, available at ptjournal.apta.org). The LEFS has potential excessive internal redundancy. Furthermore, sensitivity and long-term responsiveness are lacking, and the five-point Likert scale may increase respondent burden and scoring error. The FAAM originally was proposed and developed as a region-specific, patient-reported outcome measure, but subsequent studies used only participants with disorders below the knee. Furthermore, these studies on the FAAM show conflicting results for internal consistency. It is either demonstrated as excessive or not demonstrated sufficiently. These aspects call into question the FAAM’s ability to accurately measure lower-limb functional loss as a single kinetic unit. Consequently, the LEFS remains the only suitable criterion measure for lower-limb functional measurement.

The literature indicates that development of new lower-limb measurement tools should be considered. To be widely accepted and used, any new tool should improve the clinimetric properties of advocated patient-reported outcome measures. Psychometrics should be robust, assess function, and measure change over time. Practicality should improve readability, provide a user-friendly format, and minimize response errors through efficient completion and scoring processes. General characteristics should clarify that distribution is normalized without floor or ceiling effects and that constructs represent both function and quality of life. The objective of this study was to overcome recognized deficiencies in existing lower-limb, region-specific, patient-reported outcome measures through: (1) development of a new lower-extremity outcome scale, the Lower Limb Functional Index (LLFI), and (2) evaluation of the clinimetric properties of the LLFI using the LEFS as a criterion measure.

Materials and Method

A prospective observational study of the development and clinimetric assessment of the LLFI (Appendix 1) was completed in 2 phases (Figure). Phase 1 developed the LLFI using a 3-stage process, and phase 2 validated the LLFI in patients receiving care for lower-limb musculoskeletal conditions.

Phase 1—Development of the LLFI

Development of the LLFI followed the Kirshner and Guyatt established 3-stage process.

Stage 1—item generation. The electronic databases PubMed, CINAHL, EMBASE, and PEDro were reviewed from 1980 to 2009 with the key words “outcomes,” “self-report,” “function,” “disability,” “impairment,” “lower limb,” “leg,” “knee,” “hip,” “foot,” and “ankle.” This review identified 130 patient-reported outcome measures (eAppendix 2, available at ptjournal.apta.org). A 4-person peer panel (occupational therapist, physical therapist, general practitioner, and occupational physician) used consensus opinion, which required consensus of a minimum of 3 panel members, to review and shorten the list to 34 patient-reported outcome measures, with 873 items for lower-limb and general musculoskeletal injuries. The 873 items were further reduced to 421 items through binning and winnowing methods, which removed duplicate and non-applicable items.

Stage 2—item reduction. The peer panel further reduced the 421 items in 5 substages (2a–e). Substage 2a reduced the list to 203 items through item pooling (eg, “stairs,” “steps-up,” “steps-down,” and “slopes” became “stairs”). Substage 2b classified items using the World Health Organization International Classification of Functioning, Disability and Health (ICF) codes: b=body functions, s=body structures, d=activities and participation, and e=environmental factors. Substage 2c reduced the 203 items to 90 by combining the ICF codes to create common descriptive construct titles (eg, “stairs,” “ladders,” and “curbs” became “code
Phase 1: Development and Pilot Study (n=18)

**Stage 1: Item Generation**
- Literature search identified 130 PROs with 873 items

**Stage 2: Item Reduction**
- Removal of duplicate and non-lower limb–specific items = 160 items
- Condensing similar items = 90 items
- Patient focus group and panel feedback created LLFI questionnaire format and final 25 items

**Stage 3: Field Testing and Pilot Study**
- LLFI and LEFS completed (n=18) with responses (nR=54)

Reliability
- n=18, nR=36

Responsiveness
- n=18, nR=36

Phase 2: Validation Study (n=127)

**LLFI and LEFS**
- Australian (n=9) and American (n=1) physical therapy clinics (n=127, nR=332)

Reliability
- n=56, nR=112

Responsiveness
- n=111, nR=222

Baseline scores (n=127) provided:
- Factor structure, internal consistency, and construct validity

Pooled responses (nR=332) provided:
- Criterion validity and missing responses

Figure.
Flow chart of Lower Limb Functional Index (LLFI) development and validation. PRO = patient-reported outcome measure, LEFS = Lower Extremity Functional Scale, n = number of participants, nR = number of responses.

Lower Limb Functional Index Development and Validation

The LLFI format was based on usability for human-system interaction and user-centered design. Text boxing was used, which places questions within larger boxes to improve reader acceptability, and shadowing of alternate lines. The 3-item response option scale of “Yes,” “Partly,” and “No” was selected to provide stable, equally spaced responses.
Stage 3—field testing. A pilot investigation \( (n=18, \text{total number of responses } [n^R]=54) \) used outpatients with lower-limb conditions for preliminary LLFI assessment (Tab. 1). This pilot investigation demonstrated reliability\(^{49,49} (n=18, n^R=36) \) was high (intraclass correlation coefficient [ICC (2,1)]=.97, with 95% confidence interval [CI]), responsiveness\(^{50,51} (n=18, n^R=36) \) was high (effect size=1.3, standard response mean >1.80), and no floor or ceiling effects were present. Sampling methodology was confirmed as suitable and enabled sample size calculations for the larger main study. It also identified the patient-reported outcome measures’ change characteristic as heterogeneous,\(^{52} \) with the change coefficient demonstrated by the individual patients who had true change that varied by different amounts over 2 points in time.\(^{53} \)

Sample size. Minimum sample sizes for the validation study were calculated from the pilot study results, with an 80% likelihood of detecting differences and allowing for 15% attrition with \( P<.05.\)?\(^{54,55} \) Power calculations indicated the need for a total sample of \( n=120 \) (reliability, \( n=55 \); responsiveness, \( n=99 \); and concurrent criterion validity, \( n=104)\).?\(^{56,57} \) Exploratory factor analysis indicated a single-factor structure was likely; therefore, more than 100 participants were required.\(^{54,58} \)

Phase 2—Validation of the LLFI in a Cohort Population

A prospective, cohort design was used.\(^{26} \) Participants with lower-limb musculoskeletal conditions (\( n=127, n^R=332 \)) were recruited consecutively from 9 Australian and 1 American physical therapy clinics between 2003 and 2009. Inclusion criteria were medical practitioner referral and musculoskeletal lower-limb symptoms (acute, subacute, and chronic) that enabled a broad range of participants and conditions (Tabs. 1 and 2). Exclusion criteria, defined and determined by the referring medical practitioner and the participating therapist were: pregnancy; less than 18 years of age; English comprehension difficulty; and “red flag” signs indicating nonmusculoskeletal symptoms and lower-limb conditions, including peripheral arterial occlusive disease, deep vein thrombosis, septic arthritis, and cellulitis.\(^{59} \) In total, 142 participants were referred, with 15 being excluded (8 declined to participate, 1 was pregnant, 2 had English difficulty, 2 were excluded due to age, and 2 were referred for spinal conditions). Of the 127 participants, 111 received repeated measures, with 16 assessed only at baseline. Duration of symptom status was classified as: acute at 0 to 6 weeks, subacute at >6 to 12 weeks, and chronic at >12 weeks.\(^{60} \) Each participant’s injury was classified by region and subregion (Tabs. 1 and 2) to determine proportional representation.\(^{19,55} \)

Procedure

At initial evaluation, the participants completed the LEFS and LLFI and two 11-point external criterion clinical change scales: a global numeric rating scale (NRS) of perceived present overall status\(^{61,62} \) and a patient-specific index (PSI)\(^{26} \) that generated a list of 5 items the individual has difficulty doing. Those participants who received treatment were re-measured every 2 weeks for 6 weeks and then every 4 weeks until discharge. The LEFS is a single-page, 20-item patient-reported outcome survey questionnaire with a 5-point (0–4) Likert scale in a matrix format.\(^{22} \) The raw score is computed by totaling the points ranging from 0 to 80 (80=no disability) and multiplying the total points by 1.25 to provide a score of 0% to 100%. Up to 2 missing responses are permitted. The LLFI is a single-page, 25-item patient-reported outcome survey questionnaire with a 3-point Likert scale of 1 point for “Yes,” 0.5 point for “Partly,” and 0 points for “No.”

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>Phase 1: Pilot Study</th>
<th>Phase 2: Validation Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of participants</td>
<td>18</td>
<td>120</td>
</tr>
<tr>
<td>No. of responses</td>
<td>54</td>
<td>332</td>
</tr>
<tr>
<td>Age (y), ( X \pm SD )</td>
<td>44.5±14.1</td>
<td>44.8±15.6</td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>38.9%</td>
<td>39.7%</td>
</tr>
<tr>
<td>Injury: duration (wk), ( X \pm SD, \text{range} )</td>
<td>5.7±10.0, 1–65</td>
<td>7.3±9.1, 1–256</td>
</tr>
<tr>
<td>Hip/groin(^a)</td>
<td>11.1%</td>
<td>11.8%</td>
</tr>
<tr>
<td>Knee</td>
<td>27.7%</td>
<td>30.7%</td>
</tr>
<tr>
<td>Ankle</td>
<td>16.7%</td>
<td>22.0%</td>
</tr>
<tr>
<td>Foot</td>
<td>5.6%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Lower leg (calf, shin)</td>
<td>22.2%</td>
<td>14.2%</td>
</tr>
<tr>
<td>Upper leg (thigh)</td>
<td>5.6%</td>
<td>5.5%</td>
</tr>
<tr>
<td>Other (whole leg, ulcers)</td>
<td>5.6%</td>
<td>9.4%</td>
</tr>
<tr>
<td>Multiple areas</td>
<td>5.6%</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

\(^a\) Subregion percentages include individuals with multiple (2 or more) affected subregions, making the total greater than 100%. A statistical difference was present in the percentage representation of the categories of: ankle, foot, lower leg, and other.

Lower Limb Functional Index Development and Validation
Lower Limb Functional Index Development and Validation

Table 2.
Validation Study Participants (n=127) by Diagnoses and Percentage Representation of Total Sample as an Indication of Generalizability*

<table>
<thead>
<tr>
<th>Subregion</th>
<th>Percentage</th>
<th>Diagnoses (Alphabetic Listing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hip/groin</td>
<td>11.8</td>
<td>A-Capsule strain</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>B-Ligaments of knee: collateral, etc</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>C-Mensical cyst</td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>D-Osteoarthritis</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>E-ORIF</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>F-Perthes disease</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>G-Soft tissue strain, groin</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>H-THR/PHR</td>
</tr>
<tr>
<td>Knee</td>
<td>30.7</td>
<td>A-ACL, both conservative and surgical</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>B-Ligaments of knee: collateral, etc</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>C-Mensical cyst</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>D-Mensical, postoperative</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>E-Mensical trauma</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>F-Nonspecific knee pain</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>G-Osteoarthritis</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>H-Patellar fracture</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>I-Patellofemoral joint pain/dysfunction</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>J-Patellofemoral joint, postoperative</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>K-Patellofemoral joint subluxation</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>L-Pes anserinus inflammation</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>M-Soft tissue injury of knee</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>N-Tibiofemoral joint</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>O-ORIF</td>
</tr>
<tr>
<td>Ankle</td>
<td>22.0</td>
<td>A-Fracture of ankle, tibia/fibula</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>B-Ligament, collateral sprain, grade 1–3</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>C-ORIF</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>D-Soft tissue injury (not affecting collateral ligament)</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>E-Syndesmosis injury</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>F-Talus trauma and ligament</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>G-Talar dome fracture</td>
</tr>
<tr>
<td>Foot</td>
<td>12.6</td>
<td>A-Calcanee trauma</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>B-Cuboid</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>C-Hallux</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>D-Heel pad pain</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>E-Jones fracture</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>F-Listrac ligament</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>G-Metatarsal fracture</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>H-Metatarsalgangal joint</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>I-Nonspecific foot pain</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>J-Peripheral neuropathy</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>K-Plantar fascia</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>L-Sesamoid bone</td>
</tr>
<tr>
<td>Lower leg (calf, shin)</td>
<td>14.2</td>
<td>A-Achilles tendon</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>B-Calf strain</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>C-Compartment syndrome</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>D-Medial tibial stress syndrome</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>E-Personal strain</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>F-Tibiofibular fracture, mid shaft</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>G-Tibialis anterior muscle strain</td>
</tr>
<tr>
<td>Upper leg (thigh)</td>
<td>5.5</td>
<td>A-Hamstring muscle strain/tear grade 1–3</td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>B-Iliotibial band</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>C-Rectus femoris and quadriceps muscles</td>
</tr>
<tr>
<td>Other (whole leg, ulcers)</td>
<td>9.4</td>
<td>A-Nonspecific leg pain</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>B-Neuroma</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>C-Ulcers</td>
</tr>
<tr>
<td>Multiple areas</td>
<td>5.5</td>
<td>Diagnoses were included above</td>
</tr>
</tbody>
</table>

*ORIF = open reduction internal fixation, THR = total hip replacement, PHR = partial hip replacement, ACL = anterior cruciate ligament, TKR = total knee replacement, PKR = partial knee replacement.

Subregion and diagnoses percentage values include individuals with multiple (2 or more) affected subregions. Consequently, totals are greater than 100%.

Points are totaled, multiplied by 4, and subtracted from 100 to provide a score of 0% to 100% (0% = maximum function). Up to 2 missing responses are permitted. The global NRS for perceived status and the PSI used anchors at 0 and 10 (0 = “most possible,” 10 = “normal/no problem”). The PSI individual scores are summed, doubled, and subtracted from 100 to give a maximum functional impairment score of 0% to 100%.26

Data Analysis
The phase 2 validation (Figure) utilized participants’ baseline responses (n = 127) to assess distribution, internal consistency, and factor structure. The total responses (n = 332) assessed floor and ceiling effects, missing responses, and criterion validity between the LLFI and LEFS. Subgroups of the baseline responses were formed for reliability (n = 56, n = 112) and responsiveness (n = 111, n = 222). For responsiveness, the time period for repeated measures was based on known group differences expected to occur with natural healing and the effects of treatment. This time period, which included the immediate postoperative and postfracture period, was classified as 2 weeks for patients with acute conditions, as 4 weeks for patients with subacute conditions, and as 6 weeks for patients with chronic conditions.19,25,63 Error values of the standard error of the measurement (SEM) and the minimal detectable change (MDC) were calculated using the responsiveness subgroup (Figure). The minimal clinically important difference (MCID) was calculated using a subgroup of responses based on the clinically important change as determined by a 2-point change on the global NRS.64

Distribution and normality were assessed by visual inspection of the baseline scores histogram and...
the one-sample Kolmogorov-Smirnov test\(^5\) cutoff at a significance level of \(P>.05\). The presence of items that represented both function and quality of life were verified from face and content validity.

**Psychometric Characteristics**

**Internal consistency** was assessed by Cronbach alpha (\(\alpha=0.1.00\)) from baseline scores, with a cutoff of \(\geq.95\) indicating redundancy.\(^6\)

**Test-retest reliability** (subsample \(n=56\)) used the ICC (2,1) with 95% CI\(^6\) comparing baseline scores with scores obtained 3 days later, prior to the next treatment.\(^26,52\) Only with scores obtained 3 days later, by Cronbach alpha (\(\alpha\)) was assessed Internal consistency of life were verified from face and content validity.

**Responsiveness** was assessed from effect size (ES)\(^6\) and standardized response mean (SRM).\(^6\) Participants with repeated measures who fulfilled the \(a\ priori\) periods of expected change based on natural healing or intervention were selected as having a known group difference\(^69\) and classified by limb subregion.

**Validity** was assessed in all 4 forms. Face and content validity were assessed through the patient focus group, panel feedback, and readability scores.\(^57\) Criterion-related validity was assessed through Pearson \(r\) coefficient for concurrently completed LLFI and LEFS responses (\(n=332\)). Construct validity compared groups that changed with groups that did not change. Change was determined from the 2 external criteria over 2 time points with the NRS score at \(\pm 20\%\) change and the PSI score at \(\pm 12.5\%\) change.\(^19\) Both were required to categorize participants as improved or deteriorated,\(^70\) with an \(a\ priori\) requirement of statistical difference between the baseline and repeated groups’ paired \(t\) tests.

**Error scores** were determined with MDC at the 90% confidence interval (MDC\(_{90}\)) from the SEM formula using the ICC.\(^70\) The MCID was calculated using the distribution-based method with the construct validity definition of change on the 11-point NRS criterion measure.\(^62,64\)

**Factor analysis** was assessed from baseline LLFI and LEFS data using maximum likelihood extraction\(^54,71\) to clarify one factor for a single summed score,\(^71-73\) with loading coefficient absolute value suppression at .40.\(^54,74\) Factor extraction had 3 \(a\ priori\) requirements: scree plot point of inflection at the second eigenvalue,\(^75\) eigenvalue cutoff \(>1.0,76\) and \(\geq 10\%\) variance.\(^54,74\) SPSS version 14.0 (SPSS Inc, Chicago, Illinois) was used, with the level of significance at \(P<.05\).

**Practical Characteristics**

**Practicality** considered 9 distinct aspects.\(^26,27,77,78\) The initial 5 aspects were: (1) self-administered, (2) applicable across a variety of conditions, (3) applicable across different severity levels, (4) relevance to defined populations, and (5) single-page length. The remaining 4 areas were determined individually through the patient focus group: (1) ease of understanding and ease of completion, assessed by an 11-point global NRS anchored at 0 for “Extremely difficult” and 10 for “Extremely easy”; (2) questionnaire completion time, the average of 3 completions timed manually following 2 minutes of familiarization; (3) scoring time, the average of 3 timed scorings, following 2 practice trials, by one therapist per clinic; and (4) missing responses, a percentage of total responses. **Readability** was determined from word-processing software\(^79,80\) with Flesch-Kincaid grade scales (range = 0–12, optimum score is <7) and Flesch reading ease (optimum score is >60%).

**Assessment of Clinimetrics**

**Clinimetric performance** was assessed from 2 established clinimetric scales: (1) the Measurement of Outcome Measures (MOM) scale, which evaluates 25 aspects in 4 categories (methodological, practical, distributional, and general) by means of 3 response options (“Yes,” “Partial,” and “No”),\(^19,26\) and (2) the Bot scale, which evaluates 12 aspects through 4 response options (“Good,” “Doubtful,” “Poor,” and “Not available”).\(^27,81\) The Bot scale cutoff criteria were adjusted in 2 categories: (1) “Time to administer” was reduced from 10 to 3 minutes, and (2) “Readability and comprehension” was quantified by the Flesch-Kincaid readability criteria.\(^19,27\)

**Role of the Funding Source**

Research support was provided by the University of the Sunshine Coast.

**Results**

**Phase 1—Development of the LLFI**

The LLFI final version determined from field testing in stage 3 is presented in the Appendix. The method for item reduction used in stage 2 is presented in the Figure.

**Phase 2—Validation of the LLFI in a Cohort Population**

Patient demographics are reported in Tables 1 and 2 for both the pilot and validation phases. No inferential statistics were used.

**General Characteristics**

Distribution and normality were demonstrated through the Kolmogorov-Smirnov test for the LLFI (\(D[127]=0.074, P=.087\)) and the LEFS (\(D[114]=0.049, P=.200\)). Both measures had identical baseline score ranges (0%–98%), suitable histograms shape, and no floor or ceiling effects on visual examination. The LLFI “Partly” option was used by 43.5% of participants at baseline and in 40.1% of all responses. Face
Lower Limb Functional Index Development and Validation

Table 3.
Clinimetric Properties of the Lower Limb Functional Index (LLFI) and the Lower Extremity Functional Scale (LEFS)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Test-Retest Reliability, ICC (2,1)</th>
<th>Internal Consistency, Cronbach Alpha</th>
<th>Error Score</th>
<th>Responsiveness</th>
<th>Missing Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEM</td>
<td>MDC&lt;sub&gt;90&lt;/sub&gt;</td>
<td>SD&lt;sub&gt;100&lt;/sub&gt;</td>
<td>ES</td>
<td>SRM</td>
</tr>
<tr>
<td>LLFI</td>
<td>.97</td>
<td>.91</td>
<td>2.84</td>
<td>6.65</td>
<td>21.86</td>
</tr>
<tr>
<td>LEFS</td>
<td>.97</td>
<td>.95</td>
<td>3.48</td>
<td>8.13</td>
<td>23.22</td>
</tr>
</tbody>
</table>

*a ICC—intraclass correlation coefficient, SEM—standard error of the measurement, MDC<sub>90</sub>—minimal detectable change (90% confidence interval), SD<sub>100</sub>—standard deviation at baseline (100% scale), ES—effect size, SRM—standardized response mean.

and content validity were demonstrated only through the development (pilot) phase, with both function and quality of life being represented.

**Psychometric Characteristics**
The methodological characteristics of internal consistency, reliability, responsiveness, and error score are summarized in Table 3. Criterion validity was high (r=.88). Construct validity was demonstrated with a statistically significant (P<.001) difference between baseline and repeated measures. Mean and score differences were comparable for both patient-reported outcome measures (Tab. 4).

Factor analysis indicated a single-factor structure for the LLFI and the LEFS, with all three a priori criteria met: suitable scree plot, eigenvalue >1.0, and variances >10% (Tab. 5). The correlation matrix Kaiser-Meyer-Olkin values were .87 for the LLFI and .93 for the LEFS, with a significant Bartlett test of sphericity (P<.001).

**Practical Characteristics**
Ease of understanding and ease of completion were not significantly different, with an average score of 8.6 for the LLFI and 8.2 for the LEFS. Completion time for the LLFI required a mean (±SD) of 131±23 seconds, 29% less than the LEFS at 184±31 seconds. Scoring time required 17±5 seconds for the LLFI and 50±19 seconds for the LEFS, which increased to 150±39 seconds with missing responses. The LLFI combined completion and scoring was 1:48±28 seconds. The LEFS combined completion and scoring was 23±50 seconds with no missing responses, but 33±70 seconds with missing responses. Missing responses for the LLFI affected 1 of 332 responses, and no questionnaires were invalid. For the LEFS, 35% of the questionnaires had missing responses and 10% were invalid. Readability for the LLFI had a grade level of 7.2 and a reading ease of 61%, and the LEFS had a comparable grade level of 7.8 but a less preferable reading ease of 51%. Clinimetric performance on the MOM scale was 96% for the LLFI and 82% for the LEFS. On the Bot scale, the LLFI score was 100% (12/12) and the LEFS score was 83% (10/12).

**Discussion**

**Main Findings**
The LLFI was demonstrated as a psychometrically sound and practical patient-reported outcome survey tool suitable for assessing lower-limb function. This prospective study used external criteria that were not retrospective or reliant on recall<sup>82,83</sup> and that allowed direct comparison between the LLFI and LEFS. The LLFI demonstrated superior or equivalent clinimetric properties, in particular the psychometric characteristics of responsiveness and error values and the improved practical characteristics of missing responses, completion and scoring times, and mildly preferred readability. These findings were supported by overall clinimetric assessment via the MOM and Bot scale scores. The sample covered a large range of conditions and symptoms, which implied the results are generalizable and representative of the broader population with lower-limb musculoskeletal conditions treated in outpatient physical therapy clinics. Furthermore, these results added to the clinimetric knowledge of the LEFS.

**General Characteristics**
The 3-stage development process enabled the LLFI face and content validity to be established and supported the previous findings for the LEFS<sup>22,84</sup> This process ensured the
LLFI was acceptable and satisfactory for both clinicians and patients. The pilot study facilitated the main validation study, as it enabled estimates for sample size and assessment of the change characteristic, which reduced the potential for errors.

The baseline distribution of LEFS and LLFI scores was similar (Tab. 6). The LLFI “Partly” response was accepted by patients and used 43% of the time. Although the selection of this response afforded fewer response options compared with the 5-point LEFS format, it provided stable, equally spaced responses with 2 advantages. First, 3 options reduce the psychological dilemma of selecting a response without increased cognitive demand, as the required options of “for,” “against,” and “intermediate” are met. Second, it allows results to be analyzed with more powerful parametric statistics, as the requirement is met for the lowest level of interval data needed with individual scores that are summed to provide a total score.

### Psychometric Characteristics

The LLFI psychometric characteristics were established and preferred to those of the LEFS. The levels of internal consistency, responsiveness, error score, and factor structure all favored the LLFI. The LEFS and LLFI were equal with respect to reliability without a preference for one scale over another, with a level similar to previously published LEFS values (ICC = .88–.94). Also, the findings for construct validity were similar for both patient-reported outcome measures where both baseline and repeated measures were comparable.

### Internal Consistency

Internal consistency values favored the LLFI (α = .91) compared with the LEFS (α = .95). This result for the LEFS’s internal consistency is similar to the findings of previous studies and was sufficiently high to indicate potential item redundancy. The responsiveness values were consistently greater for the LLFI than for the LEFS. This result may simply be a consequence of the slightly lower baseline standard deviation for the LLFI. As an observational study, other influences may have included: lower baseline severity (as change rates vary between patients with acute and chronic conditions), the variation in follow-up duration for patients with acute and subacute conditions and those with chronic conditions (as an instrument is less responsive over shorter follow-up periods), and interventions provided were at the treating therapist’s discretion. For the LEFS, the responsiveness values (SEM and ES) also were marginally lower than previously reported. This finding may be attributed to our use of all known group participants anticipated to improve through natural

### Table 5.

Factor Analysis: Variance Explained for the Lower Limb Functional Index (LLFI) and the Lower Extremity Functional Scale (LEFS)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Eigenvalues</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLFI</td>
<td></td>
<td>7.57</td>
<td>30.29</td>
<td>30.29</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.92</td>
<td>7.68</td>
<td>37.96</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1.30</td>
<td>5.22</td>
<td>43.18</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.18</td>
<td>4.73</td>
<td>47.91</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1.11</td>
<td>4.44</td>
<td>52.35</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1.08</td>
<td>4.31</td>
<td>56.66</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.01</td>
<td>4.03</td>
<td>60.68</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.95–0.21</td>
<td>39.32</td>
<td>100.00</td>
</tr>
</tbody>
</table>

#### LEFS

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Eigenvalues</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>10.68</td>
<td>53.42</td>
<td>53.42</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.04</td>
<td>10.19</td>
<td>63.61</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.21</td>
<td>6.05</td>
<td>69.67</td>
</tr>
<tr>
<td>4–20</td>
<td></td>
<td>0.83–0.06</td>
<td>30.33</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* A forced one-factor solution was used with maximum likelihood extraction, varimax rotation, and a suppression of 0.40.

### Table 6.

Baseline Mean (± SD) Scores by Subregion for the Lower Limb Functional Index (LLFI) and the Lower Extremity Functional Scale (LEFS)

<table>
<thead>
<tr>
<th>Subregion</th>
<th>LLFI</th>
<th>LEFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hip/groin</td>
<td>47.8±24.6</td>
<td>48.6±29.4</td>
</tr>
<tr>
<td>Upper leg (thigh)</td>
<td>35.4±22.5</td>
<td>27.5±19.4</td>
</tr>
<tr>
<td>Knee</td>
<td>49.4±22.1</td>
<td>51.2±24.9</td>
</tr>
<tr>
<td>Lower leg (calf, Achilles tendon, shin)</td>
<td>47.2±26.3</td>
<td>47.5±20.3</td>
</tr>
<tr>
<td>Ankle</td>
<td>48.9±23.0</td>
<td>48.8±26.9</td>
</tr>
<tr>
<td>Foot</td>
<td>36.3±27.2</td>
<td>37.3±22.4</td>
</tr>
<tr>
<td>Other (whole leg, ulcers, nerve palsy)</td>
<td>57.2±24.2</td>
<td>47.7±19.3</td>
</tr>
<tr>
<td>Multiple areas</td>
<td>33.0±14.8</td>
<td>24.5±5.4</td>
</tr>
<tr>
<td>Total average</td>
<td>48.1±24.0</td>
<td>52.3±24.7</td>
</tr>
</tbody>
</table>

* n = 127
Lower Limb Functional Index Development and Validation

healing or treatment intervention, as opposed to only participants categorized as “responders.” This may have potentially reduced error indexes, but does not distinguish between the MDC64 and minimal important change.90

The LLFI error scores (SEM = 2.84%, MDC90 = 6.63%) were preferred to those of the LEFS (SEM = 3.48%, MDC90 = 8.13%) in the current study. Interestingly, the LEFS’s MDC value in this study was lower than the MDC values reported in 4 prior studies (11.3%–12.4%, 9.0–9.9 scale points).22,87,88,91 These differences probably were due to the higher reliability in the current study and, compared with the 2 recent studies,87,88 to the fact that our patients would be expected to show greater improvement through natural healing at a faster rate due to age and injury mechanism.

Factor analysis results indicated a single-factor structure; thus, the LLFI and LEFS scores can be summed for a single score. The LEFS factor structure was determined using maximum likelihood extraction19,71 and is reported for the first time. However, there were some concerns. Specifically, the cross-loading of multiple LEFS items indicated a tendency for a multifactor structure, which supported the internal consistency results and suggests the LEFS may have item redundancy.29,89 Both the LLFI and LEFS had additional factors that accounted for substantial variance. This finding suggests item reduction may be appropriate for both patient-reported outcome scales and warrants further research.

Practical Characteristics
There were practical advantages of the LLFI over the LEFS, including reduced user burden as shown by fewer missing responses, improved readability, and shorter completion and scoring times.

The clinimetric assessment on both the MOM and Bot scales demonstrated the LLFI was preferred. The LLFI’s higher levels of responsiveness and sensitivity would enable greater efficiency for researchers and clinicians in determining whether selected intervention strategies were effective. Moreover, a smaller change score and shorter time period would be required to evaluate an intervention’s outcome.

Study Limitations and Strengths
Limitations of the study were that only participants from physical therapy outpatient clinics were included and that specific conditions or diagnostic subgroups were not investigated. The results cannot be generalized to inpatient or community settings or to other body regions. The study’s strengths are the provision of a single region-specific, lower-limb, patient-reported outcome scale that improved sensitivity and reduced clinician burden and missing responses compared with a recognized criterion standard. The sample was from multiple centers and included patients with conditions affecting each subregion of the lower limb, with varied degrees of severity and duration, who represented both the general and work-injured populations. These attributes ensure the LLFI fulfills the recommendations stipulated by previous researchers6,8,50 for any new patient-reported outcome measure to ensure it is applicable to outpatients with lower-limb disorders.

Implications for Further Research
The high correlation indicating criterion validity between the LLFI and LEFS implied the LLFI could be generalized to populations in which the LEFS had been validated.51,92 However, this generalization must be made with caution until further investigations of the LLFI in other general and diagnostically specific populations are made. Independent validation is essential, and further concurrent investigation is needed. Furthermore, investigations are warranted to assess potential item reduction to shorten the LLFI, further reducing respondent and clinician burden.32

Conclusions
The LLFI is a practical patient-reported outcome measure to assess functional status in patients with lower-limb conditions. Compared with the LEFS, the LLFI demonstrated preferred clinical utility and improved clinimetric performance due to superior psychometric and practical characteristics. These findings indicate the LLFI is a viable patient-reported outcome measure for the evaluation of lower-limb status and impairment in clinical and research settings.

Mr Gabel, Dr Burkett, and Dr Michener provided concept/idea/research design. All authors provided writing and data analysis. Mr Gabel and Dr Burkett provided data collection and project management. Mr Gabel provided participants, facilities/equipment, and clerical support. Dr Melloh provided institutional liaisons. Mr Gabel, Dr Melloh, and Dr Michener provided consultation (including review of manuscript before submission). The authors thank all participating patients, general practitioners, and therapists for their time and effort.

The study was approved by the Human Research Ethics Committee of the University of the Sunshine Coast.

This research, in part, was presented at the Scientific Meeting and Annual Conference of the American Physical Therapy Association; June 8–11, 2005; Boston, Massachusetts, and at the Fourteenth Biennial Australian Physiotherapy Association-Musculoskeletal Physiotherapy Australia Conference; November 24–26, 2005; Brisbane, Australia.

Research support was provided by the University of the Sunshine Coast.


References
Lower Limb Functional Index Development and Validation


8 Naal FD, Impellizzeri FM, Rippstein PF. Which are the most frequently used outcome instruments in studies on total ankle arthroplasty? *Clin Orthop Relat Res*. 2010;468:815–826.


Lower Limb Functional Index Development and Validation


86 Jacoby J, Matell MS. Three point Likert scales are good enough. J Mark Res. 1971;8:495–500.


Appendix.
The Lower Limb Functional Index (LLFI)

LOWER LIMB FUNCTIONAL INDEX

NAME: ________________________ INJURY ________________________ □ LEFT LEG □ RIGHT LEG

DATE: _______

PLEASE COMPLETE: Your leg/s may make it difficult to do some things you normally do. This list contains sentences people use to describe themselves with such problems. Think of yourself over the last few days.

If an item describes you, mark the line. If not, leave it blank.
If an item partly describes you - Use a Half (½) Mark.

DUE TO MY LEG/S:

____ 1. I stay at home most of the time.
____ 2. I change position frequently for comfort.
____ 3. I avoid heavy jobs (eg, cleaning, lifting more than 5 kg or 10 lb, gardening).
____ 4. I rest more often.
____ 5. I get others to do things for me.
____ 6. I have the pain/problem almost all the time.
____ 7. I have difficulty lifting and carrying (eg, shopping bags up to 5 kg or 10 lb).
____ 8. My appetite is now different.
____ 9. My walking or normal recreation or sporting activity is affected.
____ 10. I have difficulty with normal home or family duties and chores.
____ 11. I sleep less well.
____ 12. I need assistance with personal care (eg, washing, hygiene).
____ 13. My regular daily activities (work, social contacts) are affected.
____ 14. I am more irritable and/or bad tempered.
____ 15. I feel weaker and/or stiffer.
____ 16. My transport independence is affected (driving, public transport).
____ 17. I have difficulty or need help with dressing (eg, trousers/pants/shoes and socks).
____ 18. I have difficulty changing directions, twisting or turning.

(Continued)
Appendix. Continued

____ 19. I am unable to move as fast as I would wish.

____ 20. I have difficulty with prolonged or extended standing.

____ 21. I have difficulty bending, squatting, and/or reaching down.

____ 22. I have difficulty with long or extended walks.

____ 23. I have difficulty with steps and stairs.

____ 24. I have difficulty with sitting for prolonged or extended times.

____ 25. I have problems with my balance on uneven surfaces and/or with unaccustomed footwear.

**LLFI SCORE:** To score the upper part, add the marked boxes:

\[
\text{TOTAL (LLFI points above)} \times 4 = \frac{\text{TOTAL}}{100} \text{ points.}
\]

**FINAL TOTAL** \((100 - \text{TOTAL}) = \%\)

**MDC (90% CI):** 6.67% or 1.67 LLFI points. Change less than this may be due to error.

* MDC (90% CI) — minimal detectable change (90% confidence interval). The LLFI may not be used or reproduced without written permission of the authors.