

On “Lower Limb Functional Index...” Gabel CP, Melloh M, Burkett B, Michener LA. *Phys Ther.* 2012;92:98–110.

We were interested to read the article by Gabel and colleagues,¹ who conducted a head-to-head comparison study of the psychometric properties of the Lower Extremity Functional Scale (LEFS),² an instrument we developed in 1999, and the Lower Limb Functional Index (LLFI), an instrument developed by the authors. We have long been advocates of head-to-head comparisons of competing instruments to determine which has the greatest potential to positively affect clinical care.³ We would like to make some general comments regarding the conceptual framework, and then make more specific comments regarding the methods and literature interpretation in Gabel and colleagues' article.

We developed the LEFS² based on the World Health Organization's model of disability and handicap.⁴ The more contemporary terms consistent with the current version of the *International Classification of Functioning, Disability and Health* (ICF)⁵ that guided instrument development are “activity limitations” and “participation restrictions.” Because our focus was on people with musculoskeletal disorders of the lower extremity, all of the items in our scale captured the person-level activity limitations and participation restrictions most relevant to people with disorders of the lower extremity. Notably absent from the LEFS are questions related to impairments (eg, pain, joint stiffness) or mental health status (eg, irritability, depression). Our rationale for this approach was that we saw problems with other functional status instruments available

at the time because they combined questions related to impairments, such as pain and joint stiffness, with items dealing with person-level function and items related to psychological distress. An example of a scale that combines the constructs of impairment, function, and mental health is the Roland-Morris Scale, a low back pain instrument developed in the 1980s.⁶ The scale includes questions dealing with pain intensity, appetite, and irritability in addition to common functional activities. Older instruments that combined multiple constructs into a single scale make it difficult for clinicians to interpret changes that may occur following treatment. Did the change score reflect improved person-level functional status, or did the change reflect some combination of changes in pain, psychological distress, and function?

In our experience, clinicians are most interested in a patient's person-level functional status and changes in functional status following treatment. This is not to say that other attributes such as psychological distress or pain are not important. Quantifying and addressing pain and mental health often are an important aspect of physical therapy intervention, but is most effectively addressed with measures designed for these purposes. In the end, however, we are most interested in the effects of our interventions on a person's ability to perform daily life activities.

In the past decade, there has been tremendous growth in the science of outcome measure development.⁷ The clear trend in this substantial amount of work is movement away from outcome measures that include multiple constructs, or what we have referred to as hybrid measures,⁸ toward those that are con-

ceptually pure and are designed to capture a singular and well-defined construct. Work by the leaders in outcomes measure development from multiple fields of study support the effort to define and capture well-defined concepts based on, for example, the ICF.^{9,10} The Patient Reported Outcome Measurement Information System (PROMIS), for example, is an effort, funded by the National Institutes of Health, to develop conceptually clear outcome measures of well-defined health concepts such as physical function, self-efficacy, pain interference, and anger.^{11,12} This movement in outcome measure development toward conceptually clear health concepts is in sharp contrast to the LLFI developed by Gabel et al, an instrument strongly reminiscent of the Roland-Morris Scale, developed almost 3 decades ago.

The instrument developed by Gabel and colleagues was designed as a measure of lower-limb function, but contains items dealing with appetite, pain, irritability, and joint stiffness. We are concerned about the lack of conceptual clarity in this measure and the potential confusion that could arise in interpreting the meaningfulness of changes in the measure. The instrument cannot differentiate between changes due to improved person-level physical function—the primary target of physical therapy interventions—and changes attributable to different health concepts such as physical impairment, irritability, or appetite. The factor analysis reported in the article supports this contention. Gabel et al reported in Table 5 that the LLFI is shown to have 7 factors with eigenvalues greater than 1, and the first factor accounts for 30.29% of the variance. In contrast, the LEFS demonstrated 3 factors, with the first factor accounting for

53.42% of the variance. The low loading of the first factor and the large number of factors relative to the LEFS provide evidence to support our concern regarding the multiple concepts included in the new instrument.

Gabel et al state that the responsiveness also favored the LLFI. We found no formal statistical comparison of the responsiveness coefficients for the 2 measures. Just as one would not accept that one therapy was superior to another based only on descriptive comparisons of point estimates in a randomized trial, neither is it appropriate to consider the responsiveness of one measure to be superior to another without a formal statistical comparison.

With respect to interpretations of the literature, Gabel et al claimed that the “LEFS lacks sensitivity to change.”¹ In referring to the LEFS, they stated, “Furthermore, sensitivity¹³ and long-term¹⁴ responsiveness are lacking.” Referring to the references cited by Gabel et al, Watson stated, “The LEFS...demonstrated high test-retest reliability and appears to be moderately responsive to clinical change in patients with anterior knee pain.”¹³ Watson et al performed a receiver operating characteristic (ROC) curve analysis and obtained an area under the curve (AUC) of .77 (95% confidence interval=.57, .97).¹³ Lin et al also performed a ROC curve analysis and obtained an AUC of .84 (95% confidence interval=.57, 1).¹⁴ This literature provides evidence beyond our original article that supports the responsiveness of the LEFS.

We found errors in the approach that Gabel and colleagues used to score the LEFS. Gabel and colleagues stated, “The raw score [of the LEFS] 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.”¹ Both of these statements are incorrect. The LEFS is scored and interpreted on a scale of 0 to 80,² and there is no provision for converting a LEFS score to a percentage. When scoring the LEFS, up to 4 missing item responses are permitted, with a maximum of 2 from any 1 of the 4 difficulty levels.¹⁵ Both of these points are critical, as they affect Gabel and colleagues’ calculations of time to complete and score the test and percentage of invalid questionnaires.

The development of self-report outcome measures is an ever-evolving science as research methodology is developed and new scales build upon the strengths and avoid the weaknesses of older scales. As such, we support the goal of the authors to develop an improved measure of lower-extremity function that is clinically relevant and efficient to utilize. As new scales are developed, it can be extremely difficult for busy clinicians to determine whether there is a sufficient body of evidence to warrant incorporation of the new measure into their clinical practice. We feel that the concerns raised in this letter are sufficient to warrant caution on the part of physical therapists as they consider this new measure for use in clinical practice.

Jill M. Binkley, Daniel L. Riddle, Paul W. Stratford

J.M. Binkley, PT, MCISc, FAAOMPT, CLT, TurningPoint Women’s Healthcare, 11180 State Bridge Rd, Suite 503, Alpharetta, GA 30022 (USA).

D.L. Riddle, PT, PhD, FAPTA, Departments of Physical Therapy and Orthopaedic Surgery, Virginia Commonwealth University, Richmond, Virginia.

P.W. Stratford, PT, MSc, School of Rehabilitation Science, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

This letter was posted as a Rapid Response on November 17, 2011, at ptjournal.apta.org. It was posted in response to the version of the article published ahead of print on November 3, 2011.

References

- 1 Gabel CP, Melloh M, Burkett B, Michener LA. Lower Limb Functional Index: development and clinimetric properties. *Phys Ther.* 2012;92:98–110.
- 2 Binkley JM, Stratford PW, Lott SA, Riddle DL, North American Orthopaedic Rehabilitation Research Network. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther.* 1999;79:371–383.
- 3 Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* 1996;76:1109–1123.
- 4 *International Classification of Impairments, Disabilities, and Handicaps.* Geneva, Switzerland: World Health Organization; 1980.
- 5 *International Classification of Functioning, Disability and Health: ICF.* Geneva, Switzerland: World Health Organization; 2001.
- 6 Roland M, Morris R. A study of the natural history of back pain, part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine.* 1983;8:141–144.
- 7 Riddle DL, Stratford PW, Singh JA, Strand CV. Variation in outcome measures in hip and knee arthroplasty clinical trials: a proposed approach to achieving consensus. *J Rheumatol.* 2009;36:2050–2056.
- 8 Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum.* 2008;59:876–883.
- 9 Jette AM, McDonough CM, Haley SM, et al. A computer-adaptive disability instrument for lower extremity osteoarthritis research demonstrated promising breadth, precision, and reliability. *J Clinical Epidemiol.* 2009;62:807–815.
- 10 Guccione AA, Mielenz TJ, Devellis RF, et al. Development and testing of a self-report instrument to measure actions: Outpatient Physical Therapy Improvement in Movement Assessment Log (OPTIMAL). *Phys Ther.* 2005;85:515–530.
- 11 Hung M, Clegg DO, Greene T, Saltzman CL. Evaluation of the PROMIS physical function item bank in orthopaedic patients. *J Orthop Res.* 2011;29:947–953.

- 12 Patient Reported Outcomes Measurement Information System (PROMIS). Available at: <http://www.nihpromis.org>. Accessed November 10, 2011.
- 13 Watson CJ, Propps M, Ratner J, et al. Reliability and responsiveness of the Lower Extremity Functional Scale and the Anterior Knee Pain Scale in patients with anterior knee pain. *J Orthop Sports Phys Ther.* 2005;35:136–146.
- 14 Lin CW, Moseley AM, Refshauge KM, Bundy AC. The Lower Extremity Functional Scale has good clinimetric properties in people with ankle fracture. *Phys Ther.* 2009;89:580–588.
- 15 Stratford PW, Hart DL, Binkley JM, et al. Interpreting lower extremity functional status scores. *Physiother Can.* 2005;57:154–162.

[DOI: 10.2522/ptj.2012.92.1.181]

Author Response

We thank Binkley, Riddle, and Stratford for their detailed and insightful commentary¹ regarding the conceptual framework, methods, and interpretation of our study on the Lower Limb Functional Index (LLFI)² and the editors of **PTJ** for the opportunity to respond. The developers of the Lower Extremity Functional Scale (LEFS)³ have provided interesting and valuable insight into the issues of developing a patient-rated outcome tool. The use of the LEFS as the criterion standard was an appropriate choice for our study's primary aim of developing a new patient-rated scale and to evaluate the clinimetric properties.

The LLFI was developed to measure function, defined in the first paragraph of our article as “the level of activities an individual performs to realize the needs of daily living.”² Binkley and colleagues used the World Health Organization's 1948 model of disability and handicap⁴ in the development of the LEFS, whereas we used the subsequently published *International Classification of Functioning, Disability and Health* (ICF)⁵ framework to develop the LLFI, as described in the “Method” section of our ar-

ticle. This included items classified across the ICF domains of body functions and structures, activity, participation, and personal and environmental factors because these all have an impact on the disorder or disease. Therefore, we thought it was critical to include all factors that have an impact on a patient's overall ability to perform the level of activities needed for daily living and participation. We are aware of opposing views on scale development. A measure developed to capture a single construct, with a single factor, is appealing. As stated by Binkley, Riddle, and Stratford, recent research and commentaries indicate a move away from “hybrid” outcome measures that include multiple constructs in favor of those that are “conceptually pure” (eg, measuring one construct). However, this may not be ideal. The ICF indicates that environmental and personal factors interact with activity limitations and participation restrictions. An individual is a biopsychosocial being, where the body and mind are integrated and each reflects the other. Clinicians are most interested in perceived person-level functional status responses that indicate whether selected interventions have affected a person's ability to perform daily activities, but distinguishing which factor is responsible and untangling the body-mind interface may prevent this from truly being determined. The LLFI considers more than just the activity limitations and participation restrictions as recommended within the ICF.⁵

We used “maximum likelihood extraction” exploratory factor analysis to determine factor structure. Table 5 reflected factors above an eigenvalue cutoff of 1.0, an arbitrary number, as the total number of factors equals the number of

questionnaire items. It is how the question-items group within each factor and satisfaction of the 3 *a priori* criteria that are essential. Although LLFI variance on factor 1 was lower compared with that of the LEFS, factor 2 did not fill the criterion of variance >10% or screeplot inflection. The LEFS factor 2 by contrast approached these criteria. Consequently, both the LLFI and LEFS demonstrated single-factor structures and from our study can be used as single summated scores. Additional analysis should be performed with a larger sample to enable the gold standard of confirmatory factor analysis. This should be the aim for all patient-reported outcomes.

We hope that the concern by Binkley, Riddle, and Stratford regarding responsiveness was answered by Table 3, where standard response mean and effect size were reported for both tools. The marginally higher scores for the LLFI are provided. We did not statistically compare the responsiveness indices between the LLFI and LEFS and did not imply that the conclusion drawn about the preference of the LLFI as compared with the LEFS, with respect to responsiveness, was based on a statistical comparison. Future studies should make statistical comparison between the LEFS and LLFI on the responsiveness indices.

The interpretations of the lack of sensitivity to change and long-term responsiveness from prior research^{6,7} were based on the fact that these studies examined responsiveness over the very short term (2–3 days)⁷ or short to medium term.⁶ Responsiveness of the LEFS was not questioned, but rather the context of time in the prior research.^{6,7} Thank you for the opportunity to clarify this statement.