Late Life Function and Disability Instrument: I. Development and Evaluation of the Disability Component

Alan M. Jette,1 Stephen M. Haley,1 Wendy J. Coster,1 Jill T. Kooyoomjian,1 Suzette Levenson,2 Tim Heeren,2 and Jacqueline Ashba2

1Roybal Center for Enhancement of Late-Life Function, Sargent College of Health and Rehabilitation Sciences, Boston University, Massachusetts.
2School of Public Health, Boston University, Massachusetts.

Background. Efforts to evaluate the effectiveness of clinical and community-based interventions designed to impact late-life disability have been hindered significantly by limitations in current instrumentation. More conceptually sound and responsive measures of disability are needed.

Methods. Applying Nagi’s disablement model, we wrote questionnaire items that assessed disability in terms of frequency and limitation in performance of 25 life tasks. We evaluated their validity and test-retest reliability with 150 ethnically and racially diverse adults aged 60 and older who had a range of functional limitations, using factor analysis and Rasch analytic techniques to examine and refine the instrument.

Results. Our analyses resulted in a 16-item disability component with two dimensions, one focused on frequency of performance and the other addressing limitation in performance of life tasks, with two disability domains within each dimension. The frequency dimension consisted of a personal and a social role domain, and the limitation dimension consisted of an instrumental and a management role domain. Expected differences in summary scores of known-functional limitation groups support the validity of this instrument. Test-retest intraclass correlations of the reproducibility of each overall dimension summary score were moderate to high (intraclass correlation coefficients .68–.82).

Conclusions. The Late-Life Function and Disability Instrument has potential to assess meaningful concepts of disability across a wide variety of life tasks with relatively few items.
travel; exchange of information; social, community, and civic activities; home life; paid or volunteer work; and involvement in economic activities. Items were refined based on a comprehensive review of existing instruments, a review by six experts, and suggestions solicited from several focus groups of older adults.

We hypothesized that it would be important to assess the frequency of performing life tasks and the person’s limitation in his or her capability to perform each task. We hypothesized that interventions could differentially affect frequency and/or limitation in performing life tasks. Accordingly, the frequency dimension describes the individual’s regularity of participating in life tasks. Frequency questions are phrased, “how often do you do a particular task?” with response options (and prompts) of “very often” (a major part of your life), “often” (a regular part of your life), “once in a while” (an occasional part of your life), “almost never” (rarely a part of your life), and “never.” The limitation dimension describes capability of performing these life tasks. We defined limitation to include both personal (health, physical, or mental energy) and environmental (transportation, accessibility, or socioeconomic) factors. Limitation questions ask “to what extent to you feel limited in doing a particular task!” with response choices of “not at all,” “a little,” “somewhat,” “a lot,” and “completely.” We framed the disability questions in a general fashion because we were interested in including factors other than health conditions (e.g., physical environment) that might limit disability outcomes.

Sampling Procedures
We recruited by telephone a convenience sample of community-dwelling adults who were 60 years of age and older. Exclusion criteria included (i) moderate or severe cognitive impairment, (ii) hospitalization of more than 1 night within the past 6 months, or (iii) inability to lift oneself out of bed. Cognitive status was defined from scores on the Short Portable Mental Status Questionnaire (SPMSQ) (13). Errors in five (or more) out of 10 items constituted moderate to severe intellectual impairment. Scores from the 10 physical function items (PF-10) of the Short Form-36 Health Survey (SF-36) (14) were used to classify subjects into four physical functioning strata based on cut-scores designed to distribute individuals into distinct ability levels. This has been used as screening criteria in previous work (15–17). Eligible subjects were interviewed by a trained interviewer in their homes. The two components of the Late-Life FDI took, on average, 25 minutes to administer. A random subset of 15 subjects completed a second interview within 1 to 3 weeks after the initial interview to assess test-retest reliability of the disability component.

Exploratory Factor Analysis
A series of factor analyses was used to identify latent factors that could be responsible for the covariation in the disability data. The principal axis method was used for the initial extraction of factors with orthogonal rotation. A one-factor and a two-factor model were tested with comparisons of percentage of the variance explained by each factor in the two models. A maximum likelihood (ML) chi-square test was performed on the one-factor model. Cronbach alpha values were calculated for the sets of items from the one-factor and the two-factor models to confirm that the item-compositions of the retained factors were correlated (i.e., measure the same construct).
Rasch Analysis

Using WINSTEPS (18), a series of one-parameter Rasch rating scale analyses was performed to estimate item calibrations along a common disability scale (19). Separate solutions were derived for each of the two disability dimensions, frequency and limitation, as well as for subsequent domains. The Rasch rating scale model used in these analyses assumed that each disability dimension (frequency and limitation) had a similar underlying structure within its own common 1–5-point rating scale. Thus, tasks that people did not perform frequently would end up on one end of the continuum, whereas tasks that most people performed frequently would end up on the opposite end of the continuum. Similarly, tasks on which many people reported severe limitations (i.e., “extremely limited—cannot do”) and tasks on which most people reported few or no limitations would end up on opposite ends of the continuum. Item calibrations were expressed in log-odd units (logits) positioned along an interval scale (20). For ease of interpretation, we transformed this raw logit scale to a 0–100 scale. Item calibrations and person scores are aligned along the same 0–100 continuum. Person scores are based on conversion of raw score counts to each transformed 0–100 scale (21). By our convention, items that were not frequently performed and items in which most people expressed limitations were associated with scores of increasing magnitude, with scores approaching 0 indicating poor capability and infrequent performance, and scores approaching 100 characterizing good capability and frequent performance.

Dimensionality of a scale was tested by demonstrating invariance of estimated item and person measures with the help of goodness-of-fit tests. An item that misfit the model indicated the item failed to discriminate between high and low performers in a way consistent with other items on the scale. Evidence of misfit suggested that the item does not belong to that particular dimension, or the item may be ambiguous in its interpretation by respondents so that it adds little to the scaling of capability or performance. We reported the information-weighted fit statistic (infit), which is standardized to approximate a mean of zero and a standard deviation of one. The infit statistic of 2.00 was used as the criterion for deciding if an item fit statistic was unexpected.

Figure 1. A visual display of each questionnaire item and its relation to the frequency dimension and limitation dimension, as well as its relation to the domains that emerged within each dimension.
according to the predicted model. This value was used because the infit statistic approximately follows a \( t \)-distribution, and a value of 2.00 has about a .05 probability of occurrence.

**Reliability and Validity Analysis**

The short-term stability of the disability component of the Late-Life FDI was assessed by test-retest reliability intra-class correlation coefficients (22) between initial and follow-up scores provided 1 to 3 weeks later (mean = 12 days; SD 8.3). As a measure of validity, we hypothesized that severity of disability (as determined by the Late-Life FDI disability component) would be positively associated with severity of functional limitation (based on SF-36 physical functioning scale, PF-10). For each scale and dimension, multiple range tests [least significant difference test (LSD)] were performed as well as post-hoc analyses using an adjusted alpha level based on the three pair-wise comparisons of interest (Bonferroni). Using the adjusted alpha, differences in mean estimates were tested for statistical significance.

**RESULTS**

**Sample Characteristics**

Twelve of the people screened did not meet eligibility criteria, while four of those eligible did not participate due to scheduling conflicts. The resulting sample was composed of 150 community-dwelling older volunteers from central and eastern Massachusetts (Table 1). The sample’s physical functioning and mental health was comparable to the older population in the United States as reflected by PF-10 and Mental Health-5 (MH-5) values on the SF-36 (23). The mean physical functioning of the sample was 61.50 (SD 31.48), compared to a population mean of 61.22 (SD 28.12). The mental health mean of the sample was 75.07 (SD 18.30), compared to a population mean of 75.29 (SD 19.20). For the disability component of the Late-Life FDI, the overall sample had mean frequency (mean = 51.33, SD 7.02) and limitation (mean = 68.59, SD 13.13) scores (higher score indicates less disability) near the midpoint of the distribution, with minimal floor (0% frequency; 0% limitation) and ceiling (0% frequency; 6.7% limitation) effects.

**Questionnaire Content**

Based on sample feedback and initial factor analyses, seven items were deleted from further analyses. Four items were eliminated because they were not applicable to a significant number of participants; other items were removed because they were too general or redundant or were reported by subjects to be confusing to answer. The remaining 16 disability items are listed in Figure 1.

**Disability Dimensions**

Table 2 presents the factor loading estimates for both the frequency and limitation dimensions. For the frequency dimension, a maximum likelihood chi-square test showed that one factor was not sufficient to adequately explain the data \( (\chi^2 = 218.28, p < .0001) \). Although four potential factors were identified by the minimum eigenvalue of 1.0, the two-factor solution was chosen, as it explained 39.6% of the variance. We labeled the first factor a *social role* domain because it reflected the frequency of performing various social and community tasks. A second factor was labeled as a *personal role* domain because it reflected the frequency of performing various personal tasks. One item (“take part in own exercise program”) had low factor loadings in both the one- and two-factor solutions, but was retained because of the importance of the item to the overall disability concept.

<table>
<thead>
<tr>
<th>Items</th>
<th>Frequency Social Role</th>
<th>Frequency Personal Role</th>
<th>Limitation Instrumental Role</th>
<th>Limitation Management Role</th>
<th>Limitation One-Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit Friends</td>
<td>.72</td>
<td>.60</td>
<td>.74</td>
<td>.79</td>
<td>.79</td>
</tr>
<tr>
<td>Travel out of Town</td>
<td>.67</td>
<td>.66</td>
<td>.79</td>
<td>—</td>
<td>.77</td>
</tr>
<tr>
<td>Go out to Public Places</td>
<td>.64</td>
<td>.68</td>
<td>.71</td>
<td>—</td>
<td>.80</td>
</tr>
<tr>
<td>Work at a Volunteer Job</td>
<td>.66</td>
<td>.59</td>
<td>.79</td>
<td>—</td>
<td>.79</td>
</tr>
<tr>
<td>Keep in Touch with Others</td>
<td>.59</td>
<td>.46</td>
<td>—</td>
<td>.59</td>
<td>.48</td>
</tr>
<tr>
<td>Participate in Social Activities</td>
<td>.56</td>
<td>.40</td>
<td>.73</td>
<td>—</td>
<td>.68</td>
</tr>
<tr>
<td>Invite Family and Friends Into Home</td>
<td>.53</td>
<td>.59</td>
<td>—</td>
<td>.47</td>
<td>.53</td>
</tr>
<tr>
<td>Participate in Active Recreation</td>
<td>.53</td>
<td>.58</td>
<td>.74</td>
<td>—</td>
<td>.67</td>
</tr>
<tr>
<td>Provide Assistance to Others</td>
<td>.48</td>
<td>.56</td>
<td>.71</td>
<td>—</td>
<td>.79</td>
</tr>
<tr>
<td>Provide Meals</td>
<td>—</td>
<td>.43</td>
<td>.57</td>
<td>—</td>
<td>.71</td>
</tr>
<tr>
<td>Take Care of Personal Care Needs</td>
<td>—</td>
<td>.35</td>
<td>.57</td>
<td>—</td>
<td>.66</td>
</tr>
<tr>
<td>Take Care of Local Errands</td>
<td>—</td>
<td>.62</td>
<td>.65</td>
<td>—</td>
<td>.76</td>
</tr>
<tr>
<td>Take Care of Health</td>
<td>—</td>
<td>.35</td>
<td>—</td>
<td>.78</td>
<td>.41</td>
</tr>
<tr>
<td>Take Care of Household Business</td>
<td>—</td>
<td>.39</td>
<td>—</td>
<td>.69</td>
<td>.46</td>
</tr>
<tr>
<td>Take Part in an Exercise Program</td>
<td>—</td>
<td>.36</td>
<td>.66</td>
<td>—</td>
<td>.65</td>
</tr>
<tr>
<td>Take Care of Inside of Home</td>
<td>—</td>
<td>.59</td>
<td>.62</td>
<td>—</td>
<td>.71</td>
</tr>
<tr>
<td>ML Test: (H(_0): 1-factor model is sufficient)</td>
<td>ML ( \chi^2 = 218.28 ) (df = 104, ( p &lt; .0001 ))</td>
<td>ML ( \chi^2 = 201.39 ) (df = 104, ( p &lt; .0001 ))</td>
<td>39.6%</td>
<td>53.9%</td>
<td>45.9%</td>
</tr>
</tbody>
</table>

Cronbach Alphas .80 .73 .82 .92 .63 .92

| Table 2. Estimates of Factor Loading for Two- and One-Factor Models With Cronbach Alphas for Frequency and Limitation

---

**Cronbach Alphas**

- .80
- .73
- .82
- .92
- .63
- .92

---

**ML Test:** (H\(_0\): 1-factor model is sufficient) ML \( \chi^2 = 218.28 \) (df = 104, \( p < .0001 \))

**% Variance Explained by Models**

- 39.6%
- 28.0%
- 53.9%
- 45.9%

---

**RESULTS**

**Sample Characteristics**

Twelve of the people screened did not meet eligibility criteria, while four of those eligible did not participate due to scheduling conflicts. The resulting sample was composed of 150 community-dwelling older volunteers from central and eastern Massachusetts (Table 1). The sample’s physical functioning and mental health was comparable to the older population in the United States as reflected by PF-10 and Mental Health-5 (MH-5) values on the SF-36 (23). The mean physical functioning of the sample was 61.50 (SD 31.48), compared to a population mean of 61.22 (SD 28.12).
The correlation between the social role and the personal role domains based on factor analysis was $r = .432$ ($p < .01$).

For the limitation dimension, a maximum likelihood chi-square test showed that one factor was not sufficient to adequately explain the data ($\chi^2 = 201.39, p < .0001$). Two factors were identified by eigenvalues greater than 1.0. This two-factor model explained 53.9% of the variance. The composition of this dimension differed in comparison with the two-factor solution for the frequency dimension. Twelve items were loaded on a factor we labeled *instrumental role* domain that encompassed activities at home and in the community. A second factor emerged that we labeled *management role* domain because it appeared to identify items that involve organization or management of social tasks that involve minimal mobility or physical activity. The correlation between the instrumental role and management role domains based on factor analysis was $r = .567$ ($p < .01$).

Figure 1 displays the relation of each questionnaire item to the frequency and limitation dimensions and their disability domains.

**Disability Instrument Scaling**

Figures 2 and 3 depict the hierarchical orders of items on the frequency and limitation dimensions of the Late-Life FDI disability component. Items are arranged according to item calibrations indicating a continuum of frequency of participation (Figure 2) or degree of limitation (Figure 3). Item separation statistics for the frequency dimension (item separation = 9.39) and its domains [social role (item separation = 8.67) and personal role (item separation = 6.35)] as well as the limitation dimension (item separation = 7.39) and its domains [instrumental role (item separation = 7.48) and management role (item separation = 5.88)] all exceed a value of 3.0, an indicator of good spread of items along a scale.

**Instrument Reliability and Validity**

Both the overall frequency and limitation dimensions discriminated between all three adjacent categories of functional limitation as indicated in Table 3. Total frequency and limitation dimension scores were the best discrimina-

---

**One Solution**

![One Solution Diagram]

**Two Solutions**

![Two Solutions Diagram]

Figure 2. Rasch models for one and two solutions of frequency. *Identifies items with fit scores $\geq$ 2.0. See text for details.
tors; in fact, in almost all cases, even the separate domains within frequency and limitation were able to discriminate among functional limitation strata. The overall frequency dimension and the instrumental role domain of the limitation dimension discriminated best among adjacent groups.

Test-retest intraclass correlations (ICC; random model) of 15 individuals for each overall dimension summary score (Table 4) were .68–.82. These correlations were obtained by removing one individual from the analyses who was an obvious outlier. Removal of the one outlier also improved the ICC estimate of frequency-personal role. Only one other test-retest correlation was relatively low, the management role score \((r = .435)\) due primarily to the small number of items \((N = 4)\) in its scale. Reproducibility was best for the limitation total and instrumental role scores \((>.80)\).

**DISCUSSION**

Our analyses resulted in a 16-item disability measure that includes two dimensions, one focused on frequency of performing life tasks with the other addressing limitation in performing these life tasks. Given the modest overlap between the two dimensions, we included distinct summary scores for the frequency and limitation dimensions in the disability component of the Late-Life FDI. Consistent with Nagi’s conceptual framework, the items in the disability component of the Late-Life FDI include major life tasks in a wide variety of social areas and thus extend beyond the traditional focus on ADLs and IADLs (1). Items such as taking part in an exercise program, providing care or assistance to others, and traveling out of town are included alongside more traditional items, such as taking care of personal care needs, included in existing ADL and IADL instruments (1,24–26).

Our analyses revealed two role domains within the frequency and limitation dimensions. Within the frequency dimension, domains of social role and personal role were identified; within the limitation dimension, domains of instrumental role and management role emerged. These role domains are only moderately correlated with each other, thus supporting the generation of distinct domain scores.
within the frequency and limitation dimensions of the disability component. Rasch analyses confirmed the hierarchical nature of each disability dimension and role domain, confirmed the adequate spacing of items along each dimension, and confirmed that Rasch-derived mean summary scores distinguish known-disability subgroups.

We were not surprised to see differences in the item grouping of the role domains within each disability dimension. Previous studies have found that responses to self-report questions differ depending upon whether older persons responded to what they actually do versus what they are capable of doing (27, 28). In our sample, the grouping of items in the dimension of frequency seemed to be independent of mobility and physical skills, as items that require competence in physical functioning loaded on different factors ["take care of local errands" (personal) and "go out with others" (social)]. Instead, items converged around domains of personal and social role. In contrast, responses to questions about limitations in performing tasks appear to reflect the importance of physical skills, as a large factor (instrumental role) emerged that seems to represent the ability to move around the home and community. The second factor in the limitation dimension grouped items around activities involving management skills, such as communication and social planning. This management role, linked to cognitive functioning, is similar to what others have often called the cognitive or advanced ADL domain (29–31). In subsequent field testing, we will evaluate the importance of including both the frequency and limitation dimensions to evaluate change in disability status. In some individuals, meaningful change in disability status may not be detected if only one aspect is addressed. For example, a person may not perceive a change in limitation in doing a task, but performs a more difficult task less often (32).

The disability structure that emerged from this study is similar to findings reported by Avlund and her colleagues, who have developed a measure of functional ability for healthy older persons. They used two evaluative criteria, labeled dependency and tiredness. Although the evaluative criteria were different than in the current study, domains of mobility function and personal ADLs were identified with the dependency criteria (33). Our model, however, is more complex, as personal and social factors are associated with frequency of performance, and instrumental and management factors are associated with perceived limitations in capability. The three factors (mobility and lower and upper limb functioning) that emerged from the tiredness evaluation criteria were more consistent with the structure we found for physical function (12). The factor analytic results indicate that the two-factor solutions were more acceptable than the one-factor solutions. The results of the Rasch scaling analyses supported both the one- and two-factor solutions as reasonable hierarchical scales, although fit problems are more evident in the one-factor than the two-factor solutions. For practical measurement purposes, however, these minor fit problems were tolerable, and the hierarchies of both the one- and two-factor models provide a good representation of the continuum of

<table>
<thead>
<tr>
<th>Table 3. Average Disability Summary Scores and Mean Differences Between Adjacent Functional Limitation Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Means (SD)</strong></td>
</tr>
<tr>
<td><strong>Severe</strong></td>
</tr>
<tr>
<td>Limitations</td>
</tr>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td>Social role</td>
</tr>
<tr>
<td>Personal role</td>
</tr>
<tr>
<td>Limitation</td>
</tr>
<tr>
<td>Instrumental role</td>
</tr>
<tr>
<td>Management role</td>
</tr>
</tbody>
</table>

Note: The difference between the means was tested using the least significant difference t test, which controls the Type I comparisonwise error rate. The alpha level for statistical significance was adjusted for the three pairwise comparisons of interest.

*p < .0167 (alpha = 0.05/3 comparisons).

**p < .003 (alpha = 0.01/3 comparisons).

Table 4. Test-Retest Reliability

<table>
<thead>
<tr>
<th>**No.</th>
<th><strong>Mean Scores (SD)</strong></th>
<th><strong>Reliability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test 1 (SD)</strong></td>
<td><strong>Test 2 (SD)</strong></td>
<td><strong>Difference (1–2)</strong></td>
</tr>
<tr>
<td>Frequency</td>
<td>16</td>
<td>51.7 (4.6)</td>
</tr>
<tr>
<td>Social role</td>
<td>9</td>
<td>46.1 (6.5)</td>
</tr>
<tr>
<td>Personal role</td>
<td>7</td>
<td>63.7 (14.6)</td>
</tr>
<tr>
<td>Limitation</td>
<td>16</td>
<td>70.4 (13.0)</td>
</tr>
<tr>
<td>Instrumental role</td>
<td>12</td>
<td>69.8 (15.5)</td>
</tr>
<tr>
<td>Management role</td>
<td>4</td>
<td>86.9 (10.7)</td>
</tr>
</tbody>
</table>

*Intraclass correlations (ICCs) are calculated using Shrout-Fleiss formulae assuming the two time points are from a random set of all possible time points. ICCs based on removing one obvious outlier (see text for details).
disability defined by frequency of performance and limitation capability. Use of either (or both) of the total dimension and domain scores will be driven by the specific instrument application. Total dimension scores provide more precision and improved stability, whereas domain scores may be more responsive to a particular area of disability assessment. Results of the known-groups analyses confirm the item content spread of each dimension of disability and suggest that the dimensions will be able to discriminate among older persons with different levels of disability. The evidence for known-groups validity is an initial step in determining the potential of the disability component to be responsive to within-person change.

Consistent with other self-reported disability and health status measures (28,34), test-retest correlations of the total dimension scores were in moderate to high range. In this small sample, the stability of frequency of participation was slightly lower than self-reported limitation. These data propose that a person’s self-assessment of capability to perform life tasks is potentially more stable than one’s perception of frequency of performance.

In summary, the disability component of the Late-Life FDI has potential to assess change in disability across a wide variety of life tasks with relatively few items. While the structure of the disability construct is complex, preliminary field-testing suggests that the disability component can discriminate among known groups that reflect the spectrum of life tasks among community-dwelling older adults. In subsequent field tests, we will evaluate the ability of the disability component of the Late-Life FDI to detect meaningful change in disability status across dimensions and role domains in a larger sample of community-dwelling older persons and compare its operating characteristics to those of other instruments commonly used in gerontologic research.

Acknowledgments

This project was funded by the National Institutes of Health, National Institute on Aging, Grant AG11669. We thank all of the volunteers who participated in our study and New England Research Institute for recruiting 50 subjects for the pilot test of the initial instrument draft. To obtain a copy of the Late-Life FDI, please access www.bu.edu/roybal/products on the World Wide Web.

Address correspondence to Alan M. Jette, Professor and Dean, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA 02215. E-mail: ajette@bu.edu

References


Received September 11, 2001
Accepted October 25, 2001