A Critical Review of Scales of Activities of Daily Living

Mary Law, Lori Letts

Key Words: measurement scales • research design

Occupational therapists routinely perform activities of daily living (ADL) assessments. Although the literature contains many ADL scales, few sources summarize and review the measurement properties of such scales. In this paper, standard criteria are used to review scales of basic self-care. Each scale is critically appraised regarding its purpose, clinical utility, construction, standardization, reliability, and validity. Recommendations are made regarding the ADL scales that are most suitable for describing, predicting, or evaluating ADL function. This review is intended to help therapists in selecting the most appropriate ADL measure to use in their clinical practice.

Mary Law, MSc, OT(C), is Assistant Professor in the Departments of Medicine and Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, and Research Manager, Occupational Therapy, Chedoke-McMaster Hospitals, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5.

Lori Letts, OT(C), at the time this article was written, was Occupational Therapy Supervisor, Para Med Health Services, Chatham, Ontario. She is currently a staff therapist, Chedoke McMaster Hospitals, Hamilton, Ontario.

This article was accepted for publication January 20, 1989.
Examples of scales are used to illustrate the review process in each methodological area. See Table 1 for a review of 13 ADL scales for each of the six measurement domains.

### Measurement Issues in the Evaluation of ADL Scales

Numerous general methodological issues apply to the review of ADL scales. A clear understanding of these issues will help the clinician in evaluating and choosing an appropriate tool.

Over the past 40 years, many scales have been developed to measure ADL function. In fact, there is currently an overabundance of these instruments in the literature (Feinstein, Jospehy, & Wells, 1986; McDowell & Newell, 1987). The problem with so many measures is that most have not been adequately validated. Few authors have a sound theoretical basis on which to develop a measure (McDowell & Newell). Rather, most authors simply select items from existing measures, make some adjustments for their particular clinical situation, and execute limited testing of reliability and validity.

Because the patient's perspective is rarely considered in instrument construction, ADL scales may not reflect the value that patients place on activities (Feinstein et al., 1986). A survey by Chiou and Burnett (1985) demonstrated that patients and therapists may value ADL functions very differently.

ADL scales are often constructed and validated for specific age groups and diagnostic groups. These scales should be constructed with a universal disability framework rather than with a diagnostic framework (Feinstein et al., 1986; Granger, 1985; Jette, 1980; Keith, 1984; Klein & Bell, 1982). The use of a disability framework (World Health Organization, 1980) would shift the emphasis of the measure to those self-care activities required for daily function rather than to specific ADL skills affected by certain diagnoses, thus preventing the development of new scales for each diagnostic category. Scoring methods can be developed to include consideration of the applicability of each item. For example, in Klein and Bell's universal ADL scale, gender-specific items can be rated not applicable and the person's score is not penalized. The universality of such a scale makes it more useful to therapists who see patients with various medical conditions but with similar self-care dysfunction.

More specific psychometric criteria that need to be considered in reviewing ADL measures for occupational therapy are the measure's (a) purpose, (b) clinical utility, (c) construction, (d) standardization, (e) reliability, and (f) validity. We will discuss these issues in more detail and show how current ADL scales meet these criteria.

### Method

We conducted a literature search of the years 1960–1988 using MEDLINE and relevant journals to identify ADL measures. The reference lists of ADL articles and books were then used to complete the search. The measures reviewed for this paper were selected on the basis of available psychometric information. If a tool was described in only one article with no information about reliability and validity, it was not included in this review; this eliminated eight scales.

The selected ADL measures were reviewed by an occupational therapy research intern trained and supervised by the first author. The reviewer used a standard measurement review form (Law, 1987). All ratings were examined by the first author, and the intern and the first author discussed any disagreements. In the two cases where a consensus could not be reached, a third rater, a senior occupational therapist, was used and the opinion of the majority prevailed. A summary of the source and the review findings for each instrument is shown in Table 1.

### Results

**Question 1: What is the Purpose of the Scale?**

Instruments are designed to quantify information for one of three purposes: description, prediction, or evaluation. A descriptive ADL instrument presents a picture of a person's status at one moment in time (Kirshner & Guyatt, 1985); the results permit comparison with other persons. A predictive ADL instrument sets criteria against which a person's status is compared. An evaluative ADL instrument measures a person's status over time to evaluate any change in ADL function.

Because the classification of measures by their purpose is relatively new in the epidemiological literature, the ADL scales reviewed were rarely classified by their purpose. Some of the measures, however, have characteristics that would support their use for one purpose over another. For example, the Index of ADL (Katz & Akpom, 1976; Katz, Downs, Cash, & Grotz, 1970; Katz, Ford, Moskowitz, Jackson, & Jaffe,
Table 1
Results of the Activities of Daily Living (ADL) Scale Review

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Descriptive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Predictive</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Evaluative</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Clinical utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>Excellent</td>
<td>Excellent Performance</td>
<td>Excellent Performance</td>
<td>Good Performance</td>
<td>Excellent Performance</td>
<td>Excellent Performance</td>
</tr>
<tr>
<td>Format</td>
<td>Performance</td>
<td>Performance</td>
<td>Performance</td>
<td>Performance</td>
<td>Performance</td>
<td>Performance</td>
</tr>
<tr>
<td>Population</td>
<td>Adult rehabilitation</td>
<td>Adult rehabilitation</td>
<td>Adult rehabilitation</td>
<td>Stroke</td>
<td>Adult rehabilitation</td>
<td>Gerontology</td>
</tr>
<tr>
<td>Completion time</td>
<td>1-2 hr</td>
<td>1 hr</td>
<td>1 hr</td>
<td>2-3 hr</td>
<td>1-2 hr</td>
<td>1-2 hr</td>
</tr>
<tr>
<td>Scale construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item selection</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Excellent</td>
<td>Good</td>
</tr>
<tr>
<td>Weighing</td>
<td>Ordinal</td>
<td>Ordinal</td>
<td>Ordinal</td>
<td>Ordinal</td>
<td>Ordinal</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Level of measurement</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Standardization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manual</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Standardization studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal consistency</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Excellent</td>
</tr>
<tr>
<td>Observer</td>
<td>Excellent</td>
<td>Poor</td>
<td>Excellent</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
</tr>
<tr>
<td>Test-Retest</td>
<td>Poor</td>
<td>Poor</td>
<td>Excellent</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Construct</td>
<td>Poor</td>
<td>Poor</td>
<td>Excellent</td>
<td>Poor</td>
<td>Poor</td>
<td>Excellent</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Poor</td>
<td>NA</td>
<td>Good</td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Overall utility</td>
<td>Fair</td>
<td>Poor</td>
<td>Excellent</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
</tr>
</tbody>
</table>

Note: NA = not applicable. Poor = poor or not considered.

In 1963, the PULSES Profile (Granger, Albrecht, & Hamilton, 1979), and the Barthel Index (Mahoney & Barthel, 1965) can discriminate between persons with and without ADL disability and can rate the severity of the disability. As a discriminative measure, the Klein-Bell ADL Scale (Klein & Bell, 1982) accurately predicts those persons who can be discharged to a community setting versus those who need placement in an extended-care facility. The Klein-Bell scale also has the potential to be a responsive, evaluative measure because of the gradation of the ADL items. A study of the use of the Klein-Bell scale in pediatrics has shown some evidence of its responsiveness (Law & Usher, 1988). The Index of ADL, the Barthel Index, the Donaldson ADL Evaluation Form (Donaldson, Wagner, & Gresham, 1973), the Kenny Self-Care Evaluation (Schoening et al., 1965), and the revised Level of Rehabilitation Scale (LORS-II) (Carey & Posavac, 1982) also show potential as evaluative measures.

**Question 2: Is the Scale Clinically Useful?**

For therapists working in clinical settings, clinical utility is of paramount importance. Clinical utility encompasses the issues of instructions, cost, time, acceptability, and format. The instructions for a measure should be clear and concise to expedite use. The scoring system should be straightforward and quick but with results that can be applied to program planning or the evaluation of clients. The ADL measures should be inexpensive, to comply with budget restrictions. The time required to score an ADL scale is also an important consideration. Most of the ADL scales reviewed for this paper are easy to administer and score and require less than 1 hour to complete; however, the time can vary depending on the client's speed and ability. An ADL scale should also be acceptable to the client and to his or her family, who should understand and agree with the usefulness of the items being measured.

An ADL measure can vary from a self-administered form, to an interview format, to the scoring of actual performance. The validity of self-report or interview measures is controversial. Do these formats truly measure what a client is capable of doing in ADL function? Klein-Parris, Clermont-Michel, and O'Neill (1986) found that the accuracy of an interview to evaluate functional skills was improved when the clients were high-functioning and the scale items were less...
complex. McGinnis, Seward, Dufong, and Osberg (1986) compared therapists’ ratings with patients’ self-reports on the Barthel Index and found that the self-report scores were significantly lower. More research is required to determine the validity of self-report measures.

**Question 3: Was the Scale Construction Adequate?**

Instrument construction involves item selection, level of measurement, and item weighting. Appropriate item selection depends on the purpose of the measure. For example, a descriptive measure should include all aspects of the basic self-care domain that can discriminate between individuals, whereas an evaluative measure should only include those ADL items that are most responsive to change. Items can be selected on the basis of expert appraisal or on the basis of statistical evidence of their ability to discriminate, predict, or evaluate ADL function. Examples of adequate item selection on the basis of expert appraisal are the Donaldson ADL Evaluation Form, the Kenny Self-Care Evaluation, the Barthel Index, the Klein-Bell ADL Scale, and the LORS-II. The Index of ADL, which has been scaled hierarchically, is the only measure in this sample in which the selection of items was validated statistically.

Individual items in an ADL scale can be scored with one of four levels of measurement: nominal, ordinal, interval, or ratio (Nunnally, 1978). Most ADL scales use ordinal scaling, such as the Kenny Self-Care Evaluation, which ranges from dependent (0) to independent (4). The level of measurement used has implications for the type of reliability statistic to use and for the inferences one can draw from the results. For example, when using an ordinal scale, a therapist cannot assume that there is an equal distance in terms of function between scores of 1 and 2 and scores of 3 and 4. Two persons could each achieve scores of 50 on the scale but have different patterns of ADL disability.

Weighting refers to the method of assigning different contributions to a total score for each item. Each item is weighted based on its perceived or mathematically derived contribution to the overall score. The items in the Klein-Bell ADL Scale and the Barthel Index are weighted on the basis of clinicians’ judgments. Weighting is thought to improve the sensitivity of the measure, but its use is controversial, and
some feel that weighting makes no difference (Wainer, 1976).

Question 4: Is the Scale Standardized?

Standardization refers to the use of extensive research to establish the reliability and validity of a measure. This research is usually published in a manual. The manual outlines the specific testing procedures described by the authors for test administration, scoring, and interpretation. The only ADL measures with published manuals are the Klein-Bell ADL Scale, the Donaldson ADL Evaluation Form, the Lawton ADL Test (Lawton & Brody, 1969), and the LORS-II. Without manuals, therapists find it difficult to consolidate the research information about a particular ADL scale.

If the measure is descriptive, norms for comparison should be included. Many of the ADL measures reviewed did not have manuals, and none included norms. Most authors have assumed that a normal person should be able to perform 100% of the basic self-care items.

Question 5: Is the Scale Reliable?

Reliability is an estimate of the extent to which an instrument is measuring true difference in ADL function among persons. The type of reliability required depends on the purpose of the measure (Kirshner & Guyatt, 1985). For a descriptive measure, the internal consistency of the scale and observer reliability are important, whereas for an evaluative measure, both observer and test–retest reliability are important.

The appropriate statistical measure of reliability also depends on the purpose and the measurement scale used. In the ADL scales reviewed for this paper, the reliability testing that has been done is appropriate. A common problem has been the use of inappropriate statistics; the Pearson correlation coefficient, which is a measure of correlation, has been used in place of the intraclass correlation coefficient, which is a true measure of agreement.

Our first concern regarding the reliability of these ADL scales was the unavailability of test data. Five of the measures have no published reliability data (ADL Test, Burke Stroke Time-Oriented Profile, Donaldson ADL Evaluation Form, Kenny Self-Care Evaluation, and Time Care Profile), and the data for five others are limited (ADL Rating Scale, Index of ADL, LORS-II, Physical Self-Maintenance Scale, Simulated ADL Examination). Measures that have demonstrated adequate observer and test–retest reliability are the Barthel Index (Granger et al., 1979), the Klein-Bell ADL Scale (Law & Usher, 1988), and the PULSES Profile (Granger et al., 1979).

Question 6: Is the Scale Valid?

Validity is the degree to which a test measures what it is intended to measure (Nunnally, 1978; Oyster, Han- ten, & Llorens, 1987). Validity is never proven but represents a gradual accumulation of evidence to support the validity of an ADL scale. Three types of validity—content, construct, and criterion—are commonly studied.

Content validity assesses the instrument’s coverage of all possible areas to be tested in one domain. Most of the ADL scales in this review have adequate content validity.

Construct validity refers to the agreement of testing results with predetermined hypotheses (Nunnally, 1978). For example, in an evaluative measure, one of the most important constructs is that the ADL scale will measure change in ADL function when change has actually occurred. The Index of ADL and the Barthel Index have measured changes in ADL function after intervention in controlled research (Gresham, Phillips, & Labi, 1980; Katz & Akpom, 1976; O’Toole, Goldberg, & Ryan, 1985). Responsiveness to change is the most essential validation evidence required for the use of these ADL measures as evaluation tools.

Criterion validity is an appraisal of the ADL scales’ agreement with a better criterion measure, that is, another measure that has already been determined to be valid in the same domain. In reviewing ADL scales, we did not find a criterion measure of ADL against which other ADL scales could be validated. Therefore, the primary validity concerns should be content and construct validity.

Discussion and Recommendations

Several recommendations for the use of ADL scales in occupational therapy can be made from this review. First, because of the numerous scales reported in the rehabilitation literature, we recommend that no further scales be developed. Rather, research should aim to improve and validate existing scales. The weakest area of these scales on the methodological review was the evidence for validity. Further research to validate these measures is essential.

We also recommend further development of those ADL scales that appear to be of the greatest clinical use. For a descriptive, diagnostic measure, the Index of ADL, the Barthel Index, the LORS-II, and the Physical Self-Maintenance Scale (Lawton & Brody, 1969) have the best reliability and validity evidence, are short, and appear to discriminate among levels of ADL function. The Klein-Bell ADL Scale, the Index of ADL, the Kenny Self-Care Evaluation, and the Barthel Index have demonstrated evidence of the ability to predict function against other standards. For evalu-
tion, the instruments with the best potential for responsively measuring change in ADL function are the Barthel Index, the Kenny Self-Care Evaluation, and the Klein–Bell ADL Scale. We stress however, that this potential for evaluation has yet to be demonstrated.

We also recommend that occupational therapy departments consider carefully which ADL measure to use for a particular application and then to adopt that measure. The use of homemade ADL checklists should be discontinued. For example, the first author’s department has adopted the following ADL scales: the PULSES Profile or the Index of ADL for description and the Klein-Bell ADL Scale for prediction and evaluation. This procedure has enhanced communication between the therapists, other health care workers, and the patients and has minimized therapist training while improving measurement standards.

References


