17. Assessment of Outcomes Following Acquired/Traumatic Brain Injury

Katherine Salter BA, Robert Teasell MD, Jeff Jutai PhD.
# Table of Contents

17.1 Introduction .................................................................................................................. 6  
17.1.1 Evaluation Criteria for Outcome Measures ....................................................... 6  
17.2 Agitated Behavior Scale (ABS) ................................................................................. 9  
17.3 Berg Balance Scale (BBS) ....................................................................................... 12  
17.4 Community Balance and Mobility Scale (CBMS) .................................................... 15  
17.5 Community Integration Questionnaire (CIQ) ........................................................ 16  
17.6 Disability Rating Scale (DRS) .................................................................................. 22  
17.7 Fatigue Severity Scale (FSS) .................................................................................... 25  
17.8 Functional Independence Measure ........................................................................... 28  
17.9 Functional Assessment Measure (FIM+FAM) .......................................................... 33  
17.10 Galveston Orientation and Amnesia Test (GOAT) .................................................... 36  
17.11 Glasgow Coma Scale (GCS) .................................................................................. 39  
17.12 Glasgow Outcome Scale (GOS)/Extended Glasgow Coma Scale (GOSE) ........... 43  
17.13 Hospital Anxiety and Depression Scale (HADS) ..................................................... 47  
17.14 Mayo-Portland Adaptability Inventory (MPAI-4) ..................................................... 50  
17.15 Medical Outcomes Study Short Form 36 (SF-36) ................................................... 55  
17.16 Mini Mental Status Examination (MMSE) ............................................................... 62  
17.17 Neurobehavioral Functioning Inventory (NFI) ........................................................ 66  
17.18 Rancho Los Amigos Levels of Cognitive Functioning Scale (LCFS) .................. 70  
17.19 Satisfaction with Life Scale (SWLS) ....................................................................... 72  
17.20 Quality of Life after Traumatic Brain Injury (QOLIBRI) ....................................... 76  
17.21 Reference List ......................................................................................................... 80
Table Directory

Table 17.1  Selected Tools for Assessment of Outcome in ABI/TBI
Table 17.2  Evaluation Criteria and Standards
Table 17.3  Evaluation Standards – Rigor
Table 17.4  Evaluation Summary
Table 17.5  Agitated Behavior Scale
Table 17.6  Characteristics of the Agitated Behavior Scale
Table 17.7  ABS Evaluation Summary
Table 17.8  Characteristics of the Berg Balance Scale
Table 17.9  BBS Evaluation Summary
Table 17.10 Characteristics of the Community Balance and Mobility Scale
Table 17.11 Community Balance and Mobility Scale Evaluation Summary
Table 17.12 Characteristics of the Community Integration Questionnaire
Table 17.13 CIQ Evaluation Summary
Table 17.14 Characteristics of the Disability Rating Scale
Table 17.15 Disability Rating Scale Evaluation Summary
Table 17.16 Fatigue Severity Scale
Table 17.17 Characteristics of the Fatigue Severity Scale
Table 17.18 FSS Evaluation Summary
Table 17.19 Characteristics of Functional Independence Measure
Table 17.20 FIM Evaluation Summary
Table 17.21 Characteristics of the FIM+FAM
Table 17.22  FIM+FAM Evaluation Summary
Table 17.23  Characteristics of the Galveston Orientation and Amnesia Test
Table 17.24  GOAT Evaluation Summary
Table 17.25  Glasgow Coma Scale
Table 17.26  Characteristics of the Glasgow Coma Scale
Table 17.27  GCS Evaluation Summary
Table 17.28  Glasgow Outcome Scale
Table 17.29  Characteristics of the GOS and GOSE
Table 17.30  GOS/GOSE Evaluation Summary
Table 17.31  The Hospital Anxiety and Depression Scale
Table 17.32  Characteristics of the HADS
Table 17.33  HADS Evaluation Summary
Table 17.34  Characteristics of the Mayo-Portland Adaptability Inventory
Table 17.35  MPAI-4 Evaluation Summary
Table 17.36  Characteristics of the Medical Outcomes Study Short Form 36
Table 17.37  SF-36 Evaluation Summary
Table 17.38  Characteristics of the Mini Mental State Examination
Table 17.39  MMSE Evaluation Summary
Table 17.40  Characteristics of the Neurobehavioral Functioning Inventory
Table 17.41  NFI Evaluation Summary
Table 17.42  Rancho Los Amigos Levels of Cognitive Functioning
Table 17.43  Characteristics of the Rancho Los Amigos Scale
Table 17.44  LCFS Evaluation Summary
Table 17.45  Satisfaction with Life Scale
Table 17.46  Characteristics of the Satisfaction with Life Scale
Table 17.47  SWLS Evaluation Summary
Table 17.48  The QOLIBRI Scale
Table 17.49  Characteristics of the QOLIBRI
Table 17.50  QOLIBRI Evaluation Summary
17. Assessment of Outcomes Following Acquired and or Traumatic Brain Injury

17.1 Introduction
The following is a review of measurement tools used in the assessment of individuals following brain injury. The list of tools appearing here was derived by a consensus of experts working on the Evidence-Based Review of Acquired Brain Injury (ABI) literature.

The tools were chosen based upon a 3-step process. The first was the development of an inventory of current outcome measures based upon both the literature and discussions held with the rehabilitation team members who actually used the tools. The second was a consensus agreement among a panel of experts as to which was most important. Finally, there had to be sufficient research on the outcome measure in ABI populations to allow a meaningful analysis of the psychometric qualities of the tools. Those outcome measures that made it through this process were selected for review.

<table>
<thead>
<tr>
<th>Table 17.1 Selected Tools for Assessment of Outcome in ABI/TBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agitated Behavior Scale</td>
</tr>
<tr>
<td>Glasgow Outcome Scale</td>
</tr>
<tr>
<td>Berg Balance Scale</td>
</tr>
<tr>
<td>Hamilton Anxiety and Depression Scale</td>
</tr>
<tr>
<td>Community Balance &amp; Mobility Scale</td>
</tr>
<tr>
<td>Mayo-Portland Adaptability Inventory</td>
</tr>
<tr>
<td>Community Integration Questionnaire</td>
</tr>
<tr>
<td>Medical Outcomes Study SF-36</td>
</tr>
<tr>
<td>Disability Rating Scale</td>
</tr>
<tr>
<td>Mini Mental State Evaluation</td>
</tr>
<tr>
<td>Fatigue Severity Scale</td>
</tr>
<tr>
<td>Neurobehavioural Functioning Inventory</td>
</tr>
<tr>
<td>Functional Independence Measure</td>
</tr>
<tr>
<td>Rancho Los Amigos Level of Cognitive Functioning Scale</td>
</tr>
<tr>
<td>Functional Assessment Measure</td>
</tr>
<tr>
<td>Satisfaction with Life Scale</td>
</tr>
<tr>
<td>Galveston Orientation and Amnesia Test</td>
</tr>
<tr>
<td>Quality of Life after TBI</td>
</tr>
<tr>
<td>Glasgow Coma Scale</td>
</tr>
</tbody>
</table>

17.1.1 Evaluation Criteria for Outcome Measures
It is necessary to have a set of criteria to guide the selection of outcomes measures. Reliability, validity and responsiveness have widespread usage and are discussed as being essential to the evaluation of outcome measures (Duncan et al. 2000; van der Putten et al. 1999; Roberts & Counsell, 1998; Law, 2002). Finch et al. (2002) provide a good tutorial on issues for outcome measure selection.

The Health Technology Assessment (HTA) programme (Fitzpatrick et al. University of Southampton, UK, 1998) examined 413 articles focusing on methodological aspects of the use and development of patient-based outcome measures. In their report, they recommend the use of 8 evaluation criteria. Table 1 lists the criteria and gives a definition for each one. It also identifies a recommended standard for quantifying (rating) each criterion, where applicable, and how the ratings should be interpreted. The
Table 17.2 - Evaluation Criteria and Standards

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Appropriateness</td>
<td>The match of the instrument to the purpose/question under study. One must determine what information is required and what use will be made of the information gathered (Wade 1992)</td>
<td>Depends upon the specific purpose for which the measurement is intended.</td>
</tr>
<tr>
<td>2. Reliability</td>
<td>- Refers to the reproducibility and internal consistency of the instrument. - Reproducibility addresses the degree to which the score is free from random error. Test re-test &amp; inter-observer reliability both focus on this aspect of reliability and are commonly evaluated using correlation statistics including ICC, Pearson’s or Spearman’s coefficients and kappa coefficients (weighted or unweighted). - Internal consistency assesses the homogeneity of the scale items. It is generally examined using split-half reliability or Cronbach’s alpha statistics. Item-to-item and item-to-scale correlations are also accepted methods.</td>
<td>Test-retest or interobserver reliability (ICC; kappa statistics): ¹ Excellent: ( \geq 0.75 ); Adequate: 0.4 – 0.74; Poor: ( \leq 0.40 ) Note: Fitzpatrick et al. (1998) recommend a minimum test-retest reliability of 0.90 if the measure is to be used to evaluate the ongoing progress of an individual in a treatment situation. Internal consistency (split-half or Cronbach’s ( \alpha ) statistics): Excellent: ( \geq 0.80 ); Adequate: 0.70 – 0.79; Poor: &lt; 0.70³ Note: Fitzpatrick et al. (1998) caution ( \alpha ) values in excess of 0.90 may indicate redundancy. Inter-item &amp; item-to-scale correlation coefficients: - Adequate levels -- inter-item: between 0.3 and 0.9; item-to-scale: between 0.2 and 0.9³</td>
</tr>
</tbody>
</table>
| 3. Validity | Does the instrument measure what it purports to measure? Forms of validity include face, content, construct, and criterion. Concurrent, convergent or discriminative, and predictive validity are all considered to be forms of criterion validity. However, concurrent, convergent and discriminative validity all depend on the existence of a “gold standard” to provide a basis for comparison. If no gold standard exists, they represent a form of construct validity in which the relationship to another measure is hypothesized (Finch et al. 2002). | Construct/convergent and concurrent correlations: Excellent: \( \geq 0.60 \), Adequate: 0.31 - 0.59, Poor: \( \leq 0.30 \)⁴ ROC analysis – AUC: Excellent: \( \geq 0.90 \), Adequate: 0.70 – 0.89, Poor: < 0.70⁵ There are no agreed on standards by which to judge sensitivity and specificity as a validity index (Riddle & Stratford, 1999) Predictive Validity: According to Shukla et al., (2011) when using many of these instruments there is no “defined threshold score beyond which an accurate prediction can be
<table>
<thead>
<tr>
<th><strong>Criterion</strong></th>
<th><strong>Definition</strong></th>
<th><strong>Standard</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Responsiveness</td>
<td>Sensitivity to changes within patients over time (which might be indicative of therapeutic effects). Responsiveness is most commonly evaluated through correlation with other change scores, effect sizes, standardized response means, relative efficiency, sensitivity &amp; specificity of change scores and ROC analysis. Assessment of possible floor and ceiling effects is included as they indicate limits to the range of detectable change beyond which no further improvement or deterioration can be noted.</td>
<td>Sensitivity to change: Excellent: Evidence of change in expected direction using methods such as standardized effect sizes: (&lt;0.5 = \text{small};) (0.5 – 0.8 = \text{moderate}) (\geq 0.8 = \text{large}) Also, by way of standardized response means, ROC analysis of change scores (area under the curve – see above) or relative efficiency. Adequate: Evidence of moderate/less change than expected; conflicting evidence. Poor: Weak evidence based solely on p-values (statistical significance)</td>
</tr>
<tr>
<td>5. Precision</td>
<td>Number of gradations or distinctions within the measurement. E.g. Yes/no response vs. a 7-point Likert response set</td>
<td>Depends on the precision required for the purpose of the measurement (e.g., classification, evaluation, prediction).</td>
</tr>
<tr>
<td>6. Interpretability</td>
<td>How meaningful are the scores? Are there consistent definitions and classifications for results? Are there norms available for comparison?</td>
<td>Jutai &amp; Teasell (2003) point out these practical issues should not be separated from consideration of the values that underscore the selection of outcome measures. A brief assessment of practicality will accompany each summary evaluation.</td>
</tr>
<tr>
<td>7. Acceptability</td>
<td>How acceptable the scale is in terms of completion by the patient – does it represent a burden? Can the assessment be completed by proxy, if necessary?</td>
<td></td>
</tr>
<tr>
<td>8. Feasibility</td>
<td>Extent of effort, burden, expense &amp; disruption to staff/clinical care arising from the administration of the instrument.</td>
<td></td>
</tr>
</tbody>
</table>

Unless otherwise noted within the table, criteria and definitions: Fitzpatrick et al. (1998); McDowell & Newell (1996). Sources for evaluation standards: ¹Andresen (2000); Hseuh et al. (2001); Wolfe et al. (1991); ²Andresen (2000); ³Hobart et al. (2001); Fitzpatrick et al. (1998); ⁴Andresen (2000); McDowell & Newell (1996); Fitzpatrick et al. (1998); Cohen et al. 2000; ⁵McDowell & Newell (1996); ⁶Hobart et al. (2001).
Each measure reviewed was also assessed for the thoroughness with which its reliability, validity and responsiveness have been reported in the literature. Standards for evaluation of rigor were adapted from McDowell & Newell (1996) and Andresen (2000).

**Table 17.3 Evaluation Standards – Rigor**

<table>
<thead>
<tr>
<th>Thoroughness or Rigor of testing</th>
<th>Excellent – most major forms of evaluation reported.</th>
<th>Adequate – several studies and/or several types of testing reported</th>
<th>Poor – minimal information is reported and/or few studies (other than author’s)</th>
<th>N/A – no information available</th>
</tr>
</thead>
</table>

Assessments of rigor using the above standards are given along with evaluation ratings for reliability, validity and responsiveness for each measure (see below).

**Table 17.4 Evaluation Summary**

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor Results</td>
<td>Rigor Results</td>
<td>Rigor Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Floor/ ceiling</td>
</tr>
</tbody>
</table>

**NOTE**: +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver; varied (re. floor/ceiling effects; mixed results)

Ratings of +++(excellent), ++ (adequate) and + (poor) results are assigned based on the criteria and evidence presented in the standards column of Table 1. For example, If a rating of “+++” or excellent is given for validity, it means that evidence has been presented demonstrating excellent construct validity based on the standards provided and in various forms including convergent and discriminant validity as well as predictive validity.

In addition to the criteria outlined above, the following additional issues were considered:

- Has the measure been used in an ABI/TBI population?
- Has the measure been tested for use with proxy assessment?

**17.2 Agitated Behavior Scale (ABS)**

The ABS was designed to assess agitation in patients who had sustained a TBI by Allied Health Professionals (AHP) working with them (Corrigan, 1989). According to Levi et al. (2005), despite the availability of the scale, agitation remains unmeasured by most who work with the TBI population. The scale, which began as a 39 item scale, was reduced to 14 items, with each item scoring 1 to 4, (from absent to present to an extreme degree). The scale which was originally tested by nurses, occupational therapists (OT), physiotherapists (PT) and other hospital staff, was designed to be used by allied health professionals.
The total score, which is considered the best overall measure of agitation, is calculated by adding the ratings (from one to four) on each of the 14 items (see Table 17.5). The scale can also be divided into 3 subscales. *Disinhibition* subscale includes items 1, 2, 3, 6, 7, 8, 9 & 10; the *Aggression* subscale includes items 3, 4, 5 & 14; and the *Lability* subscale includes items 11, 12 & 13 (Corrigan and Bogner 1994). Individual scores of >22 on the ABS indicate high agitation, conversely scores of 21 or lower indicate low agitation (Corrigan and Mysiw, 1988).

Table 17.5 Agitated Behavior Scale (Corrigan 1989)

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Short Attention span, easy distractibility, inability to concentrate</td>
<td>_______</td>
</tr>
<tr>
<td>2.</td>
<td>Impulsive, impatient, low tolerance for pain or frustration</td>
<td>_______</td>
</tr>
<tr>
<td>3.</td>
<td>Uncooperative, resistant to care, demanding</td>
<td>_______</td>
</tr>
<tr>
<td>4.</td>
<td>Violent and or threatening violence toward people or property</td>
<td>_______</td>
</tr>
<tr>
<td>5.</td>
<td>Explosive and or unpredictable anger</td>
<td>_______</td>
</tr>
<tr>
<td>6.</td>
<td>Rocking, rubbing, moaning or other self-stimulating behaviour</td>
<td>_______</td>
</tr>
<tr>
<td>7.</td>
<td>Pulling at tubes, restraints etc</td>
<td>_______</td>
</tr>
<tr>
<td>8.</td>
<td>Wandering from treatment areas</td>
<td>_______</td>
</tr>
<tr>
<td>9.</td>
<td>Restlessness, pacing, excessive movement</td>
<td>_______</td>
</tr>
<tr>
<td>10.</td>
<td>Repetitive behaviors, motor and or verbal</td>
<td>_______</td>
</tr>
<tr>
<td>11.</td>
<td>Rapid, loud or excessive talking</td>
<td>_______</td>
</tr>
<tr>
<td>12.</td>
<td>Sudden changes of mood</td>
<td>_______</td>
</tr>
<tr>
<td>13.</td>
<td>Easily initiated or excessive crying and or laughter</td>
<td>_______</td>
</tr>
<tr>
<td>14.</td>
<td>Self-abusiveness, physical and or verbal</td>
<td>_______</td>
</tr>
</tbody>
</table>

1= absent: the behaviour is not present; 2= present to a slight degree; 3= present to a moderate degree; 4= present to an extreme degree (Corrigan, 1989).

Table 17.6 Characteristics of the Agitated Behavior Scale

| Reliability               | **Interobserver reliability**: Inter-rater correlation of the total score has been found to exceed 0.70 (Corrigan and Bogner, 1995). CC of .920 for the total score and for the subscale: Disinhibition CC of .902; Aggression CC of .900; Lability CC of .726 (Bogner et al., 1999) Results from a long term care facility also indicate good inter-rater reliability with a CC of .906 for the total score, .870 for Disinhibition, .886 for Aggression and .860 for Lability. **Internal Consistency**: Cronbach’s α scores have been consistently above .838 to .914 (Corrigan, 1989; Corrigan et al., 1995; Bogner et al., 1999). Theta scores ranged from .845-.920 indicating adequate internal consistency (Corrigan, 1989). When test the scale to determine it internal consistency with individuals living in a long term care facility, Cronbach α were .808 and .740 (Bogner et al., 1999). |
| Validity                  | **Concurrent validity**: The correlations between the ABS and the Braintree Agitation Scale GCOP and GCOD were consistently high (p<0.001) (Corrigan, 1989). **Construct Validity**: Agitation was also found to be correlated with the |
OGMS ($r=-.529, p<0.001$) and the MMS ($r=-.526, p<0.001$) (Corrigan and Mysiw, 1988).

**Predictive validity:** Improvement on various cognition scales (MMS and the FIM Cognition) have been found to predict a decrease in ABS scores (Bogner et al., 2000; Corrigan and Bogner 1995).

**Responsiveness**
Corrigan and Mysiw (1988) found that as scores on the Orientation Group Monitoring System (OGMS) and the Mini-Mental State (MMS) improved, scores on the ABS decreased. Again, as PTA was resolved and cognition scores improved, agitation scores decreased. Corrigan et al. (1996) found scores from the ABS when used to assess agitation in both an Alzheimer’s disease group and an ABI group (both young and older) there were no significant difference between the means for each group: Alzheimer’s group (23.97±3.93; older BI group (23.76±4.00) and young BI group (24.05±4.05). Further investigation found the scores on the subscale aggression differed significantly between the Alzheimer’s group and the young BI group ($p<0.036$); however there was no significant difference on the scores between the young and old BI group ($p>0.05$).

**Tested for ABI/TBI patients?**
Yes designed for and tested with a TBI population

**Other Formats**
No

**Use by Proxy?**
To be administered by hospital or community staff

**Advantages**
The scale was designed to be used specially with those who had sustained a TBI (Corrigan 1989) The ABS has also been tested with a group of individual living in a long term care facility and again showed strong internal consistency and inter-rater reliability (Bogner et al., 1999). Bogner et al. 2001 found there was a strong relationship between cognition and agitation. Higher scores on the MMS and the FIM cognitive subscales were significantly related to lower scores on the ABS (Bogner et al., 2001; Corrigan and Bogner, 1994). Administering the scale requires little time and can be completed in less than 30 minutes. Agitation is considered to be present if the score is >21 (Corrigan and Bogner, 1995). The scale is readily availability and there is no cost to use it.

**Limitations**
The ABS has yet to be validated within a wider range of rehabilitation facilities (Corrigan and Bogner, 1995). As well, one of the more significant limitations of the ABI is the risk of over diagnosis of agitation (Corrigan and Mysiw, 1988).

**Summary – ABS**
**Interpretability:** Scores on the ABS are easy to interpret: Severely agitated-$>/=36$; moderately agitated – 29 to 35; mildly agitated – 22 to 28 and no agitation $<22$ (Bogner et al., 2000)

**Acceptability:** The scale is available free of charge and requires little time for training and administration.

**Feasibility:** The ABS requires little time to complete and can be completed by all AHP working with the patient.
Table 17.7 ABS Evaluation Summary

<table>
<thead>
<tr>
<th>Rigor</th>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
<th>Floor/ceiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>+++</td>
<td>+++(IR)</td>
<td>++</td>
<td>++</td>
<td>N/A</td>
</tr>
<tr>
<td>+++(IC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver;

### 17.3 Berg Balance Scale (BBS)

The Berg Balance Scale provides a quantitative assessment of balance in older adults (Berg et al. 1989). It was intended for use in monitoring the clinical status of patients or effectiveness of treatment interventions over time (Berg et al. 1995).

The scale consists of 14 items requiring subjects to maintain positions or complete movement tasks of varying levels of difficulty. All items are common to everyday life. Administration of the scale requires a ruler, a stopwatch, chair, step or stool, room to turn 360° and 10 – 15 minutes and is administered via direct observation of task completion (Berg et al. 1995; Juneja et al. 1998). Items receive a score of 0-4 based on ability to meet the specific time and distance requirements of the test. A score of zero is represents inability to complete the item and a score of 4 represents the ability to complete the task independently. It is generally accepted that scores of less than 45 are indicative of balance impairment (Berg et al. 1992a; Zwick et al. 2000)

Table 17.8 Characteristics of the Berg Balance Scale

**Reliability**

**Test-retest:** ICC’s = 0.91 (general elderly) & 0.99 (stroke survivors) (Berg et al. 1995); ICC=0.88 (Bogle-Thorbahn et al. 1996); ICC = 0.98 (Liston & Brouwer, 1996)

**Interobserver reliability:** ICC’s = 0.92 (general elderly) & 0.98 (stroke) (Berg et al. 1995); ICC = 0.98 (Berg et al. 1992b) – Mao et al. (2002) reported an overall ICC=0.95 and a range of K_w for BBS items from 0.59 – 0.94.

**Internal Consistency:** Berg et al. (1992b) reported α = 0.96; In general elderly sample, α = 0.83 & 0.97 among stroke survivors. Item to total correlations ranged from 0.38 – 0.64 (elderly) & 0.67 – 0.95 (stroke group) (Berg et al. 1995); Mao et al. (2002) reported α = 0.92 – 0.98.

**Validity**

**Concurrent validity:** correlated with global ratings of balance provided by a carer (0.47 – 0.61), by the patients themselves (0.39 – 0.41) and TUG scores (-0.48) (Berg et al. 1992a); - with TUG scores (r=0.76; p<0.001), mobility items of the BI (r=0.67; p<0.001) and with speed and amplitude laboratory measures (Berg et al. 1992b); - Liston & Brouwer (1996) showed BBS scores related to dynamic Balance Master measures (LR-3sec LR 2-sec, FB 3 sec, FB 2 sec (all p <0.05, r≥0.45) and limit of stability movement time p<0.01 r≥0.591); Mao et al. (2002) reported strong relationships between BBS scores and Fugl-Meyer-(B) balance (r=0.90 – 0.92), postural assessment scale for stroke patients (0.92 – 0.95) at 4 assessment times (14, 30, 90 and 180 days post stroke).

**Construct validity:** scores significantly correlated in the expected direction, with BI scores (r=0.80), Fugl-Meyer scale scores (0.62 – 0.94) (Berg et al. 1992a); with BI, r= 0.86 to 0.91(Mao et al. 2002); BBS scores also reported to correlate with FIM r=0.57 to 0.70; p<0.05 (Juneja et al. 1998); r=0.76; p<0.001 (Wee et al. 1999)
Construct validity (known groups): Berg et al. (1992a, 1992b) found BBS scores differentiated groups based on use of mobility aides (p<0.0001) and location of evaluation (home, rehabilitation program, acute hospital) at the end of study follow-up (p<0.0001) (Berg et al. 1992a); Wee et al. (1999, 2003) also showed admission BBS able to discriminate groups based on discharge destination of home vs. institution (p<0.0001); BBS scores discriminative of functional subgroups, p<0.001 (Stevenson, 2001); and of groups based on ambulatory status p≤0.005 (Au-Yeung et al. 2003).

Predictive validity: handicap situation in stroke survivors 6 months post discharge (multiple regression r^2=0.66; p=0.002; Desrosiers et al. 2002); admission BBS moderately predictive of length of stay (LOS) in rehabilitation unit, r=-0.39, p<0.05; r^2 = 0.362 (Juneja et al. 1998); r=-0.36; p<0.001 when controlling for age (Wee et al. 1999); r=-0.53 controlling for age and r=-0.64 when patients were admitted to rehabilitation within 14 days of stroke (Wee et al. 2003); Wee et al. (1999) demonstrated admission BBS, age and presence of social support to be predictors of discharge destination. Admission BBS score + presence/absence of family support increased prediction accuracy re: discharge destination (Wee et al. 2003); BBS scores at 14, 30 and 90 days post stroke were predictive of motor assessment scale scores at 180 days post stroke event (Mao et al. 2002); BBS scores at admission correlated with FIM scores at discharge (r=0.56; p<0.000) and with LOS (p=0.55, p<0.000), but on regression analysis BBS score was not found to be a significant independent predictor of LOS or of total discharge FIM score (Feld et al. 2001)

Responsiveness
At 14 days post stroke event, Mao et al. (2002) report a 35% floor effect. A 28.8% ceiling effect was present at 90 days post stroke; greater relative efficiency reported for the BBS vs. BI (1.0 vs. 0.68) and larger effect size at 6 – 12 weeks post-stroke evaluation suggest less ceiling effect for BBS than BI (Wood-Dauphinee et al. 1997); Bogle Thorbahn et al. 1996 reported 11% ceiling effect. Wood-Dauphinee et al. (1997) reported an effect size of 0.66 for initial 6-week, evaluation period, 0.25 for 6 – 12 weeks and overall effect size of 0.97. Mao et al. (2002) reported significant change (p ≤0.006) between times of assessment (14, 30, 90, 180 days post stroke). Effect sizes were greatest in the interval between 14 and 30 days (0.80) and diminished the further one moved through time from the stroke event (90 – 100 days – effect size = 0.40) (Mao et al. 2002); Significant change reported from pre to post intervention testing (p<0.001) (Stevenson, 2001). Minimum discernible amount of change calculated as 5.8 (90% CI) or 6.9 (95% CI); Salbach et al. (2001) demonstrated SRM = 1.04 from 8 – 38 days post stroke – there was a significant ceiling effect (26%) noted at the 2nd evaluation.

Tested for ABI/TBI patients?* Juneja et al. 1998 (construct validity); Feld et al. 2001 (predictive validity)

Other Formats N/A

Use by Proxy? N/A

* results from studies within the population of individuals with ABI/TBI appear in italics
Advantages. The Berg Balance Scale measures a number of different aspects of balance, both static and dynamic, and does so with relatively little equipment or space required (Whitney et al. 1998; Nakamura, 1998; Zwick et al. 2000). No specialized training is required to be able to administer the BBS (Nakamura et al. 1998). The high levels of reliability reported by Berg et al. (1995) were achieved when the individuals administering the test had no specific training in the administration of the scale.

Limitations. The Berg Balance scale has been thoroughly evaluated for use among populations of individuals who have experienced stroke. At present, information regarding the reliability and validity of the BBS when used among patients with TBI/ABI is limited.

No common interpretation exists for BBS scores, their relationship to mobility status and the use of mobility aids (Wee et al. 2003). The rating scales associated with each item, while numerically identical, have different operational definitions for each number or score; a score of 2, for example, is defined differently and has a different associated level of difficulty from item to item (Kornetti et al. 2004). There is also no common score associated with successful item completion (Kornetti et al. 2004). Use of an overall score that adds together ratings with different meanings having no common reference point may not be appropriate as interpretation is difficult and very little functional information is provided about the individual patient (Kornetti et al. 2004).

A recent Rasch analysis of the Berg Balance Scale revealed that some item ratings were not used at all or were underutilized, and others were unable to distinguish between individuals with different levels of ability (Kornetti et al. 2004). Collapsing rating scales to eliminate infrequently endorsed categories and creating a common pass/fail point for each item resulted in changes to the ordering of item difficulty, reduced tendencies for ceiling effects and an improved functional definition of the 45/56 cut-off point (Kornetti et al. 2004).

Summary – Berg Balance Scale

Interpretability: There are no common standards for the interpretation of BBS scores, though there is an accepted cutoff point for the presence of balance impairment.

Acceptability: This direct observation test would not be suited for severely affected patients as it assesses only one item relative to balance while sitting. Active individuals would find it too simple. The scale is not suited for use by proxy.

Feasibility: The BBS requires no specialized training to administer and relatively little equipment or space
Table 17.9 BBS Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>Results</td>
<td>Floor/ceiling</td>
</tr>
<tr>
<td>++</td>
<td>+++ TR</td>
<td>+++</td>
</tr>
<tr>
<td>+++ (IO)</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>+++ (IC)</td>
<td></td>
<td>Varied</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver; varied (re. floor/ceiling effects; mixed results)

17.4 Community Balance and Mobility Scale (CBMS)

The Community Balance and Mobility Scale (CBMS) is a performance-based measure intended to evaluate balance and mobility skills in individuals who have experienced mild to moderate traumatic brain injury (TBI) (Inness et al. 1999). The scale is comprised of 13 items each of which are rated on a 6-point scale from 0 through 5 where 5 represent the most successful completion of the scale item (Inness et al. 1999, Butcher et al. 2004).

Table 17.10 Characteristics of the Community Balance and Mobility Scale

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest: ICC=0.975 (Inness et al., 1999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency: α=0.96 (Inness et al. 1999)</td>
<td></td>
</tr>
<tr>
<td>Validity</td>
<td>Face validity: items rated as relevant to the assessment of balance by PT's (Inness et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Construct validity: - correlated with gait variables; walking velocity (r=0.69), step length (r=0.75) and step time variability (r=-0.49); however, CB&amp;M scores did not correlate significantly with measures of postural sway or with a measure of balance confidence (ABC)</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>-SRM = 1.26 (for CB&amp;M change scores) (Inness et al. 2004)</td>
</tr>
<tr>
<td>Tested for ABI/TBI patients?</td>
<td>Developed for use in TBI population.</td>
</tr>
<tr>
<td>Other Formats</td>
<td>N/A</td>
</tr>
<tr>
<td>Use by proxy?</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Advantages.** The CBMS is a measure developed specifically for use in assessment of individuals who have sustained mild to moderate TBI. It may have increased sensitivity to change when used within this population when compared to more established measures such as the Berg Balance Scale (Inness et al. 2004).

**Limitations.** The scale may be assessing a construct more similar to “dynamic mobility” rather than balance *per se* (Inness et al. 2004). The information available in the literature with regard to the reliability, validity or practical application of this scale is extremely limited and comes from the scale authors only. Further (and broader) evaluation of the CBMS’ measurement properties is required.
Summary – CBMS

**Interpretability:** Not enough information available

**Acceptability:** Not enough information available

**Feasibility:** Not enough information available

<table>
<thead>
<tr>
<th>Table 17.1 Community Balance and Mobility Scale Evaluation Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td>Rigor</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>+++ (IC)</td>
</tr>
</tbody>
</table>

*NOTE: +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.*

### 17.5 Community Integration Questionnaire (CIQ)

The Community Integration Questionnaire (CIQ) (Willer et al. 1993) was intended as a brief assessment of community integration or the degree to which an individual post-traumatic brain injury is able to perform appropriate roles within the home and community. In order to achieve higher levels of reliability, the CIQ uses behavioural indicators of integration and does not include items focused on feelings or emotional status (Willer et al. 1994; Dijkers, 1997). The CIQ was developed for inclusion in the National Institute on Disability and Rehabilitation Research (NIDRR) TBI model systems National Data Base in the United States (Dijkers 1997).

The CIQ assesses handicap, which is viewed by the scale authors as the opposite of integration (Willer et al. 1993) in three domains: Home Integration (active participation in the operation of the home or household), Social Integration (participation in social activities outside the home) and Productivity (regular performance of work, school and/or volunteer activities). The scale is comprised of 15 items in 3 corresponding subscales each of which has a different number of items and subscores (Willer et al. 1994, Sander et al. 1999). The Home Integration subscale consists of 5 items each scored on a scale from 0 – 2, where 2 represents the greatest degree of integration. The Social Integration subscale is comprised of 6 items rated in the same manner as Home Integration whereas the Productivity subscale consists of 4 questions with responses weighted to provide a total of 7 points. Scores from each of the subscales are summed to provide an overall CIQ score. The maximum possible score is 29, which reflects complete community integration (Hall et al. 1996).

The CIQ may be administered via self-completion, face-to-face or telephone interviews (Hall et al. 1996). If the individual with TBI is unable to complete the assessment, the questionnaire may be completed by proxy (Willer et al. 1994). There are two versions of the questionnaire available, one for completion by the person with TBI and one for completion by a suitable proxy (family member, close friend, significant other) (Sander...
et al. 1999). The CIQ requires approximately 15 minutes to complete (Hall et al. 1996, Zhang et al. 2002).

Table 17.12 Characteristics of the Community Integration Questionnaire

| Reliability | Test-retest: ICC=0.86 for CIQ total, 0.88 for home integration, 0.66 for social integration and 0.80 for productivity (Cusick et al. 2000); Willer et al. (1993) reported r=0.93 for home integration, 0.86 for social integration, 0.83 for productivity and 0.91 for CIQ total; r=0.63 for productive activity, 0.70 for social integration, 0.71 for home integration and 0.81 for CIQ total scores (Seale et al. 2002); r=0.91 for patients’ and 0.97 for family members/caregivers assessment (Willer et al. 1994)
| Interobserver reliability: Willer et al. (1993) reported interrater reliability between patients with TBI and their family members of r=0.81 (home integration), 0.74 (social integration), 0.96 (productivity) and 0.89 for the total CIQ score;
| Internal consistency: \( \alpha = 0.76 \) (Willer et al. 1993); Willer et al. (1994a) reported item-to-total correlations ranging from 0.32 (socialization, with whom) to 0.67 (housework, leisure activities); \( \alpha = 0.76 \) for total CIQ, 0.84, 0.83, & 0.35 for home integration, social integration and productivity, respectively (Willer et al. 1994, cited in Dijkers, 1997); post severe TBI \( \alpha = 0.26 \) (productivity), 0.65 (social integration), 0.95 (home integration) and 0.84 (total CIQ) – subtotal to total correlations were reported to be 0.54, 0.0.74, 0.79 for productivity, social integration and home integration, respectively (Corrigan & Deming, 1995).

| Validity | Construct validity: 3 components with eigenvalues >1 were identified and maintained for orthogonal rotation – these 3 factors, labeled Home Competency, Social Integration and Productive Activity, accounted for 51% of variance in the set of variables – all items loaded significantly – finance item moved to home competency and traveling was included on social integration – shopping was excluded since it loaded significantly and equivalently on 2 factors (Sander et al. 1999); Dijkers (1997) reviewed 4 articles providing correlations between subscale scores and found moderate to weak correlations – suggesting that there are three dimensions which are related to each other and may be assessing different aspects of the same concept; Kuipers et al. (2004) reported a more stable 2 dimensional structure on multi-dimensional scaling (productivity vs. personal life & independence vs. dependence), although they also were able to identify a 3 dimensional structure in keeping with factors of home competency, social interactions and productive activities;
| Construct validity (known groups): Willer et al. (1993) reported that a group of individuals with TBI vs. a nondisabled group demonstrated significantly less integration on CIQ (total scores and all subscores) except for women who were equally integrated in the home, regardless of group membership; Differences in CIQ subscores and total CIQ scores were significant (p<0.0001) when a group of individuals with TBI and a group of non-TBI control participants were compared – Groups of patients differentiated by independent living, supported living and institutional living setting could also be distinguished by differences in CIQ scores (p<0.001, Willer et al. 1994); Corrigan & Deming (1995) reported CIQ scores did not differ significantly between groups of persons with various disabilities (2 TBI samples vs. “other disabilities”; p>0.01)
| Concurrent validity: Total CIQ scores correlated with total CHART scores (r=0.62, p<0.05) – 2 CHART subscales appear comparable to CIQ subscales (occupation & social integration) – CHART occupation correlated with all CIQ subscales and most strongly with CIQ productivity (r=0.55), CHART social integration correlated with...
CIQ \( (r=0.35) \), however, this was not sufficient to reach significance \( (p>0.05) \) (Willer et al. 1993); CIQ subscale and overall scores correlated significantly and in the expected direction with DRS items and FIM+FAM items; DRS Level of functioning scores correlated most strongly with home competency \( (-0.46) \) and total CIQ scores \( (-0.47) \), DRS Employability with CIQ productive activity \( (-0.58) \) and CIQ total scores \( (-0.58) \), FAM community access with home competency \( (0.46) \) and CIQ total \( (0.47) \), FIM social interaction with all CIQ subscales \( (0.24 \text{ – } 0.27) \) and CIQ total \( (0.34) \) and FAM Employability with CIQ productive activity \( (0.57) \) and CIQ total \( (0.60) \) (Sander et al. 1999); CIQ total scores correlated significantly with DRS total scores \( (r=-0.43, p<0.01) \), CIQ home integration correlated with DRS cognitive ability, level of function and employability subscales, CIQ social interaction and productivity scales did not correlate significantly with any of the DRS subscales. CIQ total correlated significantly with CHART totals \( (r=0.68, p<0.01) \), CHART physical correlated significantly with CIQ home integration \( (r=0.53, p<0.01) \) and social integration \( (r=0.25; p<0.05) \), CHART social interaction correlated with CIQ social integration \( (r=0.38; p<0.01) \), CHART motor correlated significantly with all CIQ subscales \( (r=0.40 \text{ – } 0.47, p<0.01) \), as did CHART occupation subscale \( (r=0.33 \text{ – } 0.42, p<0.01) \); Zhang et al. 2002); CIQ subscores correlated with ratings of ADL’s \( (r=0.37, 0.37 \text{ and } 0.40 \text{ for home integration, social integration and productivity, respectively; Heinemann & Whiteneck, 1995) } \)

**Predictive validity:** Heinemann & Whiteneck (1995) reported that Social Integration and Productivity subscale scores were the two most powerful predictors of life satisfaction on multiple regression \( (\beta=0.25 \text{ and } -0.22, \text{ respectively}) \) such that greater satisfaction was associated with less social and productive handicap.

## Responsiveness

Willer et al. (1993) reported that only 1 individual obtained a maximum CIQ score on social integration, while 10/16 obtained maximum scores on the CHART social integration subscale; To examine possible ceiling effects, CIQ scores were compared to average scores on each subscale obtained by nondisabled individuals approximately ½ of individuals with TBI reached this level 2 years post injury on the home and social interaction subscales of the CIQ while only 19% reached the average level of non TBI individuals on the productivity subscale (Hall et al. 1996); Gurkha et al. (1999) report scores at 6 months and 24 months post rehabilitation discharge to be normally distributed – CIQ sensitive to a range of levels of community integration; 20.8% of subjects obtained maximum scores on social integration, while 39.1% obtained minimum scores on productive activity one year following injury (Sander et al. 1999); Corrigan & Deming (1995) reported relatively normal distributions for CIQ totals as well as for the home integration and social integration subscales; however, the productivity subscale appeared to be positively skewed with highly restricted variability in TBI and “other disability” samples; Seale et al. (2002) reported patients receiving post-acute rehabilitation improved significantly from admission to follow-up on all CIQ indicators; patients receiving rehabilitation less than 1 year post-injury improved more than patients receiving rehabilitation more than 1 year post injury \( (F=35.82, p<0.0001, \text{ over time } r^2=0.57 \text{ vs. } F=12.95, p<0.001, \text{ over time } r^2 = 0.25) \); Willer et al. (1999) reported significant improvement of CIQ scores in treatment vs. control groups from time 1 to time 2 assessments \( (p<0.001) \) – similar improvements compared to the control group were reported for home integration, social integration and productivity; Corrigan & Deming (1995) reported significant differences \( (p<0.01) \) in CIQ scores from premorbid/retrospective ratings to follow-up/current ratings with follow-up ratings being lower than premorbid for CIQ total, social integration and

---

**Evidence-Based Review of Moderate to Severe Acquired Brain Injury**

2013

---

**Module 17 - Assessment of Outcomes Following Acquired/Traumatic Brain Injury - V9-2013**

www.abiebr.com

Updated August 2013
productivity scores – only home integration did not differ significantly from premorbid to follow-up ratings.

<table>
<thead>
<tr>
<th>Tested for ABI/TBI patients?</th>
<th>Developed specifically for individuals with TBI.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other Formats</strong></td>
<td><strong>Revised subscale &amp; scoring:</strong> Sander et al. (1999) repeated factor analysis resulting in a slightly modified subscale structure. Recommendations for a revised scale and scoring are provided. Using the revised scoring proposed by Sander et al. (1999), CIQ total scores were significantly related to CIPI social activity and inactivity subscales (r=-0.43 &amp; -0.68 respectively; p&lt;0.05) as were CIQ Home Integration (r=-0.36 &amp; -0.38; p&lt;0.05) and CIQ Social Integration (r=-0.46 &amp; -0.38; p&lt;0.05; Kaplan, 2001)</td>
</tr>
<tr>
<td><strong>Mail Administration:</strong></td>
<td>Using a mail questionnaire based on the modifications of Sander et al. (1999), Kuipers et al. (2004) reported a 80.2% completion rate for CIQ questionnaires by patients and 77.7% among proxy recipients. Home competency subscales had the highest completion rates in both groups, while social interaction had the lowest. Proxy scores on the home integration scale were significantly lower than patient scores (p=0.019). Item-to-total correlations ranged from 0.19 to 0.63 and subscale-to-total correlations were reported to be 0.73 (home integration), 0.64 (social interaction) and 0.54 (productive activities). CIQ scores correlated with scores on the Sydney Psychosocial Re-integration Scale (SPRS) (0.56 &amp; 0.60 for patient &amp; proxy scores, respectively) – SPRS Independent Living correlated with CIQ home competency (0.42 &amp; 0.57 for patient &amp; proxy respectively), CIQ Social Interaction with SPRS Interpersonal Relationships (0.45 &amp; 0.49 for patient &amp; proxy), CIQ Productive Activity and SPRS Occupational Activity (0.42 &amp; 0.41 for patient &amp; proxy scores)</td>
</tr>
<tr>
<td><strong>Use by proxy?</strong></td>
<td>Agreement between scores derived from patient vs. significant other telephone interviews was reported to be ICC=0.43 for home integration, 0.65 for social integration and 0.81 for productivity subscales of the CIQ (Tepper et al. 1996). Agreement between patient and proxy scores obtained via telephone interview was reported to be 0.78 for CIQ total, 0.79 for home integration, 0.52 for social integration and 0.84 for productivity subscales – poorest agreements were noted for items that were most subjective and required opinion/judgement. In cognitive areas, proxies tended to score patients lower than the patients did themselves, while in activity areas, proxies tended to score patients higher than the patients themselves (Cusick et al. 2000). Agreement between patient and proxy ranged from $\kappa=0.43 – 0.70$ on CIQ home integration subscale, 0.42 – 0.60 on the social integration subscale and 0.69 – 0.94 on the productivity subscale. Significant differences were reported between patient and family member ratings on the home integration subscale (p&lt;0.01) and total CIQ scores (p&lt;0.05). In both cases, patient scores indicated higher levels of integration than scores derived from family member interviews, however, 80% of variance in total CIQ scores could be attributed to home integration subscores (Sander et al. 1997). When informants were interviewed, Willer et al. (1993) reported test retest reliability of 0.97 for CIQ total scores, and r=0.90 (social integration), 0.96 (home integration) and 0.97 for productivity subscales. Correlations between ratings provided by individuals with brain injury and family members were reported to be 0.81 for home integration, 0.74 for social integration and 0.96 for productivity –</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Advantages. The CIQ has become one of the most widely used tools in the assessment of community integration for people who have experienced traumatic brain injury. The scale was originally developed via an expert panel that included individuals with TBI suggesting that items have face validity (Willier et al. 1993, 1994). The scale can be completed quickly and easily by most individuals with TBI or by an appropriate proxy. The scale focuses more on behaviour than emotional states, which promotes better agreement between patient and proxy ratings (Dijkers 1997; Cusick et al. 2000).

Limitations. While the CIQ was developed to assess handicap as defined under the ICDH, the CIQ does not appear to assess all of the domains included in the WHO definition of handicap (Dijkers, 1997). Indeed, under the current definitions provided by the International Classification of Functioning, Disability and Health (WHO, 2001), CIQ items may reflect activities more than participation (Kuipers et al. 2004). The reduction of items from 47 to 15 based on factor analysis excluded items not loading onto one of the three predetermined factors that might have provided a more comprehensive assessment of handicap and/or participation. It should be noted that the factor analysis used to eliminate scale items was based on scale scores from an extremely small sample (n=49) of individuals with severe TBI (Willer et al. 1993, Dijkers, 1997).

The CIQ does not measure integration skills, the success of integration activities from the point of view of the individual with TBI or the feelings or meaning association with integration activities (Willer et al. 1993; Zhang et al. 2002). What the CIQ does measure appears to somewhat inconsistent. Some items measure the frequency with which activities are performed, while others measure, in a limited way, the assistance or supervision required in order to perform an activity (Dijkers 1997, Zhang et al. 2002). In addition, the CIQ social integration subscale does not relate to other measures of social integration in the expected way. The CIQ social integration subscale appears inconsistently related to the CHART social interaction subscale (Willer et al. 1993, Zhang et al. 2002) and only weakly related to FIM social interaction item (Sanders et al. 1999). It has been suggested that all three may be measuring slightly difference constructs. The FIM examines appropriateness of interaction while CHART assesses the size and composition of social networks. The CIQ does not assess either of these aspects of social integration (Sanders et al. 1999).

Age, gender and level of education have all been reported to have an effect on CIQ scores. Dijkers (1997) reviewed 4 studies that reported the effects of age and, in general, it appeared as though scores for women indicated greater integration into the home, while male scores typically suggested more integration into the productivity
domain. A more recent study (Kaplan, 2001) demonstrated similar effects of gender around home integration in a sample of individuals with malignant brain tumours. It has been suggested that a lack of more traditional, male household tasks may account for some of the reported differences in home integration (Dijkers 1997). The CIQ separates the activities of running a household from productive activity and, therefore, may penalize individuals who were and continue to be homemakers as well as those individuals with family members who have always shared in home-making activities (Kaplan, 2001). It has been suggested that this bias could be ameliorated by conducting a retrospective, pre-morbid assessment to provide a basis for comparison (Sander et al. 1999).

In his 1997 review, Dijkers reported a tendency for younger age to be associated with greater integration on the CIQ. Kaplan (2001) reported that older age was significantly related to poorer community integration both for the total CIQ and for each subscale. In addition to age and gender, amount of education appears to have an effect on community integration as assessed by the CIQ such that more education is associated with better integration in all three dimensions (Heinemann & Whiteneck 1995; Kaplan, 2001). The roles of men and women, people of different ages and varying levels of education are all different and these differences need to be reflected in the scale through the development of age-appropriate norms stratified by education, gender and marital status (Dijkers 1997; Sander et al. 1999, Kaplan 2001).

In a 1999 assessment of the factor structure and validity of the CIQ, Sander et al. identified 2 items that appeared problematic. It was recommended that the childcare item and the frequency of shopping item both be removed. The childcare item is frequently not applicable and appears to penalize people who have no children in the home while the shopping item loaded significantly on 2 of the 3 identified factors and did not contribute any unique information to the sale (Sander et al. 1999).

**Summary – CIQ**

**Interpretability:** The CIQ is widely used. However, no norms are currently available. There is no basis for determining that an individual’s level of integration on the CIQ is or is not normal (Dijkers 1997).

**Acceptability:** The scale is short and simple and represents little patient burden. It has been used successfully with proxy respondents.

**Feasibility:** No special training is required to administer the CIQ. The scale is free, but should be requested from the scale author. It has been used in longitudinal studies to show change over time.
Table 17.13 CIQ Evaluation Summary

<table>
<thead>
<tr>
<th></th>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
<td>Results</td>
</tr>
<tr>
<td>++</td>
<td>++(TR)</td>
<td>++(IO)</td>
<td>++(IC)</td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.

17.6 Disability Rating Scale (DRS)
The Disability Rating Scale was developed to provide quantitative information regarding the progress of individuals with severe head injury from “coma to community” (Rappaport et al. 1982). The DRS was designed to reflect changes in the following areas: arousal and awareness, cognitive ability to deal with problems around self care, degree of physical dependence and psychosocial adaptability as reflected in the ability to do useful work (Rappaport et al. 1982). The DRS was developed and tested in a rehabilitation setting among individuals who had experienced moderate to severe traumatic brain injury (Hall, 1997).

The DRS is comprised of eight items in four categories; arousal, awareness and responsivity, cognitive ability for self-care activities, dependence on others and psychosocial adaptability (Rappaport et al. 1982). Each item has its own rating scale ranging from 0 – 3 to 0 – 5 and are either in ½ or 1-point increments. (Rating forms are available for download at [http://tbims.org/combi/drs/drsrat.htm](http://tbims.org/combi/drs/drsrat.htm)) The total or composite score is calculated by summing the ratings for all 8 items. Lower scores are associated with less disability. The overall score can be used to assign the individual to one of 10 disability outcome categories ranging from no disability (DRS score = 0) to extreme vegetative state (DRS score = 29) and death (DRS = 30) (Hall et al. 1996, Fleming and Maas,1994).

The DRS is available at no cost and is free to copy. It may be downloaded in *.pdf format from [http://tbims.org/combi](http://tbims.org/combi). Training materials are also provided on the same website and a training video is available for a modest fee. Administration of the scale may be via direct observation or interview (Hall et al. 1993). When necessary, collateral sources of information may be used to complete the ratings (Rappaport et al. 1982). The DRS is simple to administer and requires approximately 5 minutes to complete (Hall, 1997, Hall et al. 1993).

Table 17.14 Characteristics of the Disability Rating Scale

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest: r=0.95 (Gouvier et al., 1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interobserver reliability: - inter-rater correlations ranged from 0.97 – 0.98 (p&lt;0.01; Rappaport et al. 1982); average r=0.98 (Gouvier et al., 1987); correlations between observer ratings ranged from 0.75 – 0.89 (Fleming &amp; Maas 1994)</td>
<td></td>
</tr>
</tbody>
</table>
| Internal consistency: item to item correlations ranged from 0.23 to 0.95, item-to-total correlations ranged from 0.54 (eye opening) to 0.96 (feeding) (Rappaport et
### Validity

**Construct validity:** DRS ratings at admission correlated with evoked brain potential abnormality scores ($r=0.78; p<0.01$; Rappaport et al. 1978); correlations between DRS and scores in cognitive testing in the following domains – intellectual, executive, academic and visuoperceptual – ranged from -0.17 to -0.37 ($p<0.05$) suggesting that better levels of function as assessed by the DRS is associated with better performance in a given cognitive domain (Neese et al. 2000).

**Construct validity (known groups):** DRS could discriminate between groups of patients who had received cognitive rehabilitation or not (Fryer & Haffey, 1987)

**Concurrent validity:** Admission DRS scores correlated with initial Stover & Zeiger (S-Z) ratings ($r=0.92$), discharge DRS scores correlated with discharge SZ scores ($r=0.81$), GOS scores (0.80) and EGOS scores (0.85) (Gouvier et al. 1987); DRS ratings were significantly correlated with FIM motor, FIM cognition, FIM+FAM motor and FIM+FAM cognition scores ($r=0.641, 0.728, 0.680, 0.746$, respectively, all $p<0.05$), DRS rating also correlated with LCFS ratings ($r=0.708$; Hall et al. 1993); GOS scores correlated with DRS at admission to ($r=0.50, p<0.01$) and discharge from rehabilitation ($r=0.67, p<0.01$; Hall et al. 1985)

**Predictive validity:** Initial DRS scores correlated with discharge SZ scores (0.65), GOS scores (0.62) and expanded GOS scores (0.73); DRS scores at admission to and discharge from rehabilitation were both significantly related to employment status at one year post-injury (Cifu et al. 1997); initial DRS ratings correlated with DRS ratings at 12 months post-injury ($r=0.53; p<0.01$) (Rappaport et al. 1982); initial DRS score correlated with length of hospital stay ($r=0.50, p<0.01$) and with discharge DRS scores ($r=0.66, p<0.01$; Eliason & Topp, 1984); via growth curve modeling, flatter rates of recovery on the DRS recovery curve were associated with higher rates of reported cognitive difficulties, as well as severity of affective/neurobehavioural disturbance and severity and burden of physical dependence at 6 months post-injury as reported by significant others (McCauley et al. 2001); initial DRS score and rate of recovery accounted for 62% variance in discharge DRS scores ($p<0.00$; Fleming & Maas. 1994); Fryer & Haffey (1987) reported DRS at admission to rehabilitation was significantly predictive of need for supervision and return to work 1 year post injury ($r=0.77, p<0.001$); initial and discharge DRS scores significantly related to vocational status ($p<0.007$; Rao & Kilgore, 1997)

**Responsiveness**

Ceiling effects reported such that DRS scores do not discriminate effectively among patients scoring in the upper categories of the Extended Glasgow Outcome Scale (Wilson et al. 2000); Rasch analysis demonstrated that a wide range of difficulty is reflected in scale items from very simple functioning to very complex with less sensitivity at the high end (Hall et al., 1993); DRS had a 6% ceiling effect at discharge, 47% at 1-year post injury and 54% at year 2, when ceiling effect is defined as scoring in the top 10% of the scale (Hall et al. 1996)

From admission through discharge and follow-up, DRS scores rated by family members demonstrated significant change over time ($p<0.0001$)

With level in disability decreasing over the duration of rehabilitation and from rehabilitation discharge to follow-up at 3 months post-discharge (Novack et al. 1991); significant differences reported between DRS ratings at discharge from rehabilitation and one-year follow-up ($p<0.001$) (Hammond et al. 2001); from admission to discharge from rehabilitation, improvement shown by the DRS was significantly greater than that shown by the GOS (71% vs. 33%, $p<0.01$; Hall et al. 1985).
Advantages The DRS is a single assessment comprised of items spanning all major dimensions of the ICIDH (impairment, disability and handicap) (Rappaport et al. 1982, Hall et al. 1996). It is a brief and simple tool that allows for the ongoing assessment of recovery from injury to community re-integration. In addition, the ability to assign scores to outcome category with relatively little loss of information (Gouvier et al. 1987) provides a quick snapshot of the individual’s overall disability status (Hall et al. 1993). The DRS appears to be more reliable and valid than the Level of Cognitive Functioning Scale (LCFS) and may be more sensitive to change than categorical rankings such as the Glasgow Outcome Scale (Hall et al. 1985). In addition, Glasgow Coma scores can be obtained from the DRS (Hall, 1997).

Limitations Descriptions of what corresponds to successful item performance at each rating level are not precise and subscales do not clearly identify areas for intervention (Brazil, 1992). The sequellae of head injury that are included for assessment are limited and do not include cognition (Brazil, 1992). The DRS assesses only general rather than specific function or functional change (Hall and Johnston 1994). It may be most useful as a means to characterize sample severity and provide the means for comparison to other groups, but it is not particularly sensitive to the effects of treatments designed to ameliorate specific functional limitations or social participation (Hall et al. 1991). In inpatient rehabilitation settings, the FIM is a more sensitive instrument with which to monitor change (Hall and Johnston, 1994).

The DRS is not well suited to patients with mild TBI or very severe impairments (Hall et al. 1993, Hall et al. 1996, Wilson et al. 2000). It has been recommended that ½ point scoring increments rather than whole points should be employed in order to increase the precision and sensitivity of the instrument when assessing higher functioning individuals (Hall et al. 1993). When subjects do not fit whole-point definitions for cognitive ability for self-care items, dependence on others and employability, ½ points can be awarded; total scores with ½ points are rounded down for the purposes of assignment to outcome category (Hammond et al. 2001). The rating form available for download has included the ½ point scoring option. When using the ½ point scoring option, the DRS does appear to be sensitive to change between discharge and one-year and even 5-year follow-ups; however, year by year change is not captured by DRS ratings more than one year post-injury (Hammond et al. 2001).
Summary – DRS

Interpretability: The DRS is widely used and is part of the TBI Model Systems Database. It provides a quick, accessible snapshot of outcome of disability in terms of general function.

Acceptability: The simplicity and brevity associated with the DRS would suggest little to no patient burden associated with its administration. Ratings provided by family members are strongly correlated with those completed by healthcare team members.

Feasibility: The DRS is free to use and copy. Training materials are also provided free of charge and a training video is available for a modest fee. The DRS seems to be able to detect significant change over time and may be well suited for group comparisons.

Table 17.15 Disability Rating Scale Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+</td>
<td>+++(TR)</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>+++(IO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ (IC)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver

17.7 Fatigue Severity Scale (FSS)

Fatigue is essentially a subjective experience and often hard to measure even though it can be a major source of disablement (Dittner et al., 2004; Belmont et al., 2006). Individuals who sustain a TBI, regardless of the level of injury, often report fatigue as a constant or recurrent problem post injury (Borgaro et al., 2005; Belmont et al., 2006). Zinno and Ponsford (2006) found activities that required mental or physical effort often resulted in increased levels of fatigue.

Fatigue Severity Scale (FSS), is a self-report questionnaire designed to assess disabling fatigue in all individuals (Krupp et al., 1989). The scale was designed to look fatigue/function measures; that is the connection between fatigue intensity and functional disability Dittner et al., 2004; Taylor et al., 2000). The FSS, which consists of 9 questions, uses a 7 point likert scale ranging from strongly disagrees to strongly agree (see below). The scores from each question are totalled with lower scores indicating less fatigue in everyday life. The total score for the FSS is calculated as the average of the individual item responses. Although the FSS was originally designed to assess fatigue in individuals with multiple sclerosis (MS), it has been found to be sensitive to fatigue in those with a TBI (Zinno and Ponsford, 2005).
Table 17.1 Fatigue Severity Scale (FSS) (Krupp et al., 1989).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Completely Disagree</th>
<th>Completely Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My motivation is lower when I am fatigued</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>2. Exercise brings on my fatigued</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>3. I am easily fatigued</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>4. Fatigue interferes with my physical functioning</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>5. Fatigue causes frequent problems for me</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>6. My fatigue prevents sustained physical functioning</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>7. Fatigue interferes with carrying out certain duties and responsibilities</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>8. Fatigue is among my 3 disabling symptoms</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>9. Fatigue interferes with my work, family or social life</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>

Individuals are instructed to choose a number from 1 to 7 that indicates their degree of agreement with each statement where 1 indicated strongly disagree and 7 strongly agree.

Table 17.17 Characteristics of the FSS

| Reliability | Test-Retest: Test retest reliability of the scale found no significant differences on the FSS scores from time one to time two. Patients were tested at 2 time periods separated by 5 to 33 weeks (Krupp et al., 1989). When tested with a group of patients who had been diagnosed with Hep C the ICC scores were 0.82 I (Taylor et al., 2000). ICC values for Turkish study were found to be 0.81 (Armutlu et al. 2007). The scale has been found to have good test-retest reliability (Dittner et al., 2004).

Internal Consistency: Cronbach alpha scores for the FSS were .81 for an MS population, .88 for a normal healthy population. (Krupp et al., 1989). Cronbach \( \alpha \) scores for those with Hep C was 0.94 and the CC were 0.82 (Taylor et al., 2000). Paired t-tests were completed looking at the scored from the screening test and the baseline tests, but no difference were found (mean difference -0.03, t=0.95, p=0.34) (Taylor et al., 2000). Armutlu et al.2007 found Cronbach \( \alpha \) scores ranged between 0.8899 and 0.9401. Ziino and Ponsford (2005) found good internal consistency when the scale was used with a group of TBI patients (Cronbach \( \alpha \) score .90 with item total correlation ranging from .37 to .84). Overall the scale has been found to have high internal consistency (Dittner et al., 2004). It has been suggested that the scale could be shortened as there appear to be a high level of redundancy with in the scale (Amtmann et al., 2012). In a study with polio patients, Cronbach \( \alpha \) score was greater than .95, with item to total correlation ranging from .68 to .88 (Burger et al., 2010).

Validity | Concurrent validity: FSS scores have been found to be highly correlated with both VAS scores (r=-0.76 and the SF-36 (r=-0.76) (Taylor et al., 2000). Ziino and Ponsford (2005) have found the FSS,
VAS-f subscales and COF subscales were all significantly correlated. Between the COF–ME and the COF–PE a strong positive correlation was found (r=.56 for each measure). For the VAS-F (Vigour and Fatigue) lower vigor scores were associated with higher fatigue scores (Ziino and Ponsford, 2005).

**Construct validity:** LaChapelle et al. (1998) noted negative correlations between time since injury and the FSS (r=-0.42, P<0.001), the impact of fatigue on cognitive and physical functioning (r=-0.41, p<0.001; r=-0.48, p<0.48 respectively). Amtmann et al (2012), in a study that included only MS patients, found a high correlation with both the subscales of the MFIS and the MFIS total score. The FSS had the highest correlation with the MFIS-physical subscale (p=0.77) and the lowest correlation was with the MFIS Cognitive (p=0.55)

**Predictive validity:** The scale has been shown to discriminate between fatigued and non fatigued patients (Friedman et al., 2010; LaChapelle et al., 1998; Krupp et al., 1989; Taylor et al., 2000). Burger et al. (2010) found only a moderate correlation between the 3 VAS scores (daily life, self care, and household and occupation) and the FSS scores possible due to the FSS measuring only physical symptoms of fatigue compared to the VAS.

### Responsiveness
The FSS has been found to be sensitive to change with time and treatment (Dittner et al., 2004). When compared to the Modified Fatigue Impact Scale (MFIS) the FSS had floor to ceiling responses ranging from 0.9 to 6.8, while the MFIS had a range of 1.1 to 0.7 (Amtmann et al., 2012).

<table>
<thead>
<tr>
<th>Tested for ABI/TBI patients?*</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Formats</td>
<td>Currently there are two more versions of the FSS: FSS-7, FSS-5. The scale has been translated into: Australian English, Canadian English, French, Canadian French, German, Swill, New Zealand English, UK English, Mexican Spanish, Spanish and Taiwanese (Kleinman et al., 2000).</td>
</tr>
<tr>
<td>Use by Proxy?</td>
<td>No</td>
</tr>
</tbody>
</table>

### Advantages
The FSS scale is a self-report scale that is easy to administer and can be completed quickly with minimal effort (LaChapelle et al., 1998; Burger et al., 2010).

### Limitations
Although the overall score of the FSS is beneficial in determining between groups, the individual questions are not able to do so (LaChapelle et al., 1998). Because no two fatigue scales measure the same thing it is strongly recommended that the user know: what aspect of fatigue they want to assess and why; whether or not a unidimensional or multidimensional scale should be used; and final is this scale beneficial to the population being assessed (Dittner et al., 2004). Another major concern with the scale is the use of a 7 point likert scale (completely disagree to completely agree). It is believed that 6 or more categories on any rating scale obscure the distinction between the categories. The collapsing of the options to three (disagree, neutral, agree) may improve
the measure (Burger et al., 2010). The FSS has not been found to be a good instrument for measuring cognitive levels of fatigue (Amtmann et al., 2012).

Summary – FSS

**Interpretability:** The FSS has been shown to be a valid and reliable scale for several populations including the ABI population (Ziino and Ponsford, 2005). Never the less the scores on the FSS are easy to interpret and are used to assess patients for fatigue post injury. Items on the scale can be open to interpretation as the word fatigue may mean something different to each individual (Burger et al., 2010)

**Acceptability:** The scale has been shown to be both valid and reliable with a variety of populations. It has been shown to have good internal consistency and is sensitive to change in fatigue levels over time.

**Feasibility:** The FSS is a self-administered scale that does not require any training to use and is available in several languages.

<table>
<thead>
<tr>
<th>Table 17.18 FSS Evaluation Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver;

17.8 Functional Independence Measure

Developed in 1987, in part as a response to criticism of the Barthel Index, the FIM was intended to address issues of sensitivity and comprehensiveness as well as provide a uniform measurement system for disability for use in the medical remuneration system in the United States (McDowell & Newell, 1996). Rather than independence or dependence, the FIM assesses physical and cognitive disability in terms of burden of care – that is, the FIM score is intended to represent the burden of caring for that individual.

The FIM is a composite measure consisting of 18 items assessing 6 areas of function (self-care, sphincter control, mobility, locomotion, communication and social cognition). These fall into 2 basic domains; physical (13 items) and cognitive (5 items). The 13 physical items are based on those found on the Barthel Index, while the cognitive items are intended to assess social interaction, problem-solving and memory. The physical items are collectively referred to as the motor-FIM while the remaining 5 items are referred to as the cognitive-FIM. Further the scale has not been found to fit with the Rasch model with MS patients (Mills et al., 2009)

Each item is scored on a 7-point Likert scale indicative of the amount of assistance required to perform each item (1=total assistance, 7 = total independence). A simple summed score of 18 – 126 is obtained where 18 represents complete dependence/total
assistance and 126 represents complete independence. Subscale scores for the physical and cognitive domains may also be used and may yield more useful information than combining them into a single FIM score (Linacre et al. 1994).

Administration of the FIM requires training and certification. The most common approach to administration is direct observation. The FIM takes approximately 30 minutes to administer and score. The developers of the FIM further recommend that the rating be derived by consensus opinion of a multi-disciplinary team after a period of observation.

Table 17.19 Characteristics of Functional Independence Measure

| Reliability | Inter rater Reliability: In a review of 11 studies, Ottenbacher et al., 1996 reported a mean inter-observer reliability value of 0.95; a median test-retest reliability of 0.95 and a median equivalence reliability (across versions) of 0.92. Reliability was higher for items in the motor domain than for those in the social/cognitive domain; ICC = 0.98 for total FIM, 0.95 and 0.89 for motor FIM and cognitive FIM, respectively (Hobart et al. 2001); ICC = 0.85 for total FIM, 0.92 for motor FIM and 0.69 for cognitive FIM (Donaghy & Wass, 1998) |
| Construct validity: Linacre et al (1994) reported 2 distinct aspects of disability within the FIM – motor & cognitive function; Cavanagh et al. (2000) suggested simple 2-factor model of the FIM not be sufficient to describe disability following stroke (66% of variance)- may not adequately measure within patient change whereas a 3-factor model (self-care, cognition and elimination) accounted for more variance (74.2%); use of Rasch transformed scores for comparison of level of ability at the end of treatment show the motor FIM to be a discriminative, ordinal, outcome measure of disability (Brock et al. 2002; Linacre et al. 1994) |
| Validity | Construct validity (known groups): FIM scores discriminated between groups based on spinal cord injury severity (p<0.05) and presence of comorbid illness (p<0.005) and right or left –sided involvement in stroke patients both at admission (p<0.005) and discharge (p< 0.05) – most of this score difference occurred on the communication domain (Dodds et al. 1993); on admission and discharge, FIM scores discriminated groups with and without neglect (p<0.001; p<0.02) and with or without aphasia (p<0.01; p<0.09) (Ring et al. 1997) |
| Concurrent validity: motor-FIM showed strong concurrent validity in association with BI – Spearman’s correlation coefficient ranging from 0.74 (admission) to 0.92 (discharge) (Hsueh et al., 2002); Kwon et al. (2004) reported r=0.95 between motor-FIM and BI scores and –0.89 between motor-FIM and MRS scores; FIM motor scores and cognition scores were significantly correlated with DRS ratings (r=0.641 and 0.728 respectively, p<0.05) and FIM cognition scores correlated with LCFS scores (r=0.645, p<0.05; Hall et al. 1993) |
| Convergent/Discriminant validity: FIM total and motor FIM scores correlated more strongly with OPCS disability scores, LHS scores, SF-36 physical component scores and WAIS – verbal IQ, than with measures of mental health status or |
psychological distress (SF36 mental component, General Health Questionnaire), however, cognitive FIM correlated most strongly with OPCS Disability scores and WAIS-verbal IQ scores and weakly with LHS, SF36 physical and mental components, and the General Health Questionnaire (Hobart et al. 2001)

**Predictive validity:** FIM admission score predictive of placement after discharge (Oczkowski et al. 1993; Dodd et al. 1993); FIM scores + length of stay predictive of functional gain, p<0.0002 (Ring et al. 1997); Granger et al (1993) reported FIM predictive of burden of care assessed in help in minutes/day (p=0.01); Singh et al. (2000) reported FIM scores at 1 month post stroke predictive of depression at 3 months post stroke as part of predictive model that also included “living at home” and “damage to inferior frontal region”; FIM scores at admission to rehabilitation significantly associated with employment status one year post head injury (Cifu et al. 1997); admission motor FIM accounted for 52% of variance in discharge motor function among TBI patients, admission cognitive FIM scores accounted for 46% of variance in discharge cognitive function – admission motor FIM was the most significant predictor of length of stay (Heinemann et al. 1994)

**Responsiveness**

Changes in FIM scores from admission to discharge in the expected direction (p<0.0005) (Dodds et al. 1993); significant differences in FIM total, FIM motor and FIM cognition scores were reported between rehab discharge and follow-up one year post injury (p<0.0001 for all) – change between 1 & 2 years and 1 and 5 years was reported to be distributed across all items with most change in cognitive function (Hammond et al. 2001)

When ceiling effect is defined as the top 29% of the scale (scoring ≥108), 49% of TBI patients scored in this range at rehabilitation discharge, and 84% by year one post injury (Hall et al. 1996); 4% of patients obtained maximum FIM scores (McPherson & Pentland, 1997); Neither floor nor ceiling effects were reported at admission to or discharge from rehabilitation post stroke (Dromerick et al. 2003); 16% ceiling effect reported for motor FIM (Brock et al. 2002); van der Putten et. al.(1999)- no significant floor or ceiling effects when administering the FIM to stroke patients - effect sizes of 0.30, 0.34 and 0 were reported for the total-FIM, motor-FIM and cognitive-FIM respectively.; Wallace et al. (2002) reported ES = 0.28 (0.46 in known changers), SRM = 0.62 (0.94 among known changers) and AUC ROC curve = 0.675; Dromerick et al. (2003) reported SRM=2.18 from admission to discharge from rehabilitation – the FIM detected change in 91/95 individuals including change in 18 patients in whom the BI detected no change (p<0.001); FIM motor predictive of direct assistance required while FIM cognition scores predictive of amount of supervision required (Corrigan et al. 1997); SRM = 0.48 reported for FIM total and 0.54 and 0.17 for motor and cognitive FIM respectively – no significant floor/ceiling effects were reported although there was a 16.1% ceiling effect noted for cognitive FIM (Hobart et al. 2001)

**Tested for ABI/TBI patients?**

The FIM has been tested with TBI populations and with a mixed population (ABI/TBI and surgical patients) (Dodds et al. 1993; Linacre et al. 1994; Cifu et al. 1997; Hall et al. 1996; Hall et al. 1993; Heinemann et al. 1994; Hammond et al. 2001; Corrigan et al. 1997; McPherson & Pentland, 1997; Donaghy & Wass, 1998; Hobart et al. 2001)

**Other Formats**

**Standardized Interview:** Davie et al. (2001) examined the FIM home interview for intrarater stability - kappa values > 0.40 on 17/18 items; motor FIM reliability reported higher than social/cognitive items (K= 0.46 to 0.61). On sequentially separate interviews, self-care items showed K values of 0.4 – 0.6; transfers, locomotion & social/cognitive items were below 0.4 (poor).
**Telephone Interview:** Smith et al. (1996) reported total-FIM ICC = 0.97, motor-FIM ICC= 0.98, and cognitive-FIM ICC = 0.57 comparing telephone interview with direct observation in the home. Item level agreement was superior for items in the motor domain (Kappa values exceeded 0.45).

Petrella et al. (2002) reported good predictive validity (discharge FIM vs phone FIM at 8 weeks; r=0.436; p=0.02) though not as good as observed FIM scores (r=0.699, p<0.0001). Phone FIM showed good concurrent validity with the observed FIM (r=0.741, p<0.0001) and was sensitive to change over time (t=3.603, p=0.001).

Duncan et al. (2002) reported a 46% ceiling effect on the motor FIM when administered by telephone at 6-month follow-up of stroke patients.

**Use by Proxy?**

Segal et al. (1996) - ICC's for patient assessment vs. proxy assessment for both in-person and telephone interviews of 0.90 & 0.91, respectively - agreement was much higher for motor-FIM than for cognitive-FIM. The authors speculate that, due to the more subjective nature of the cognitive dimension of the FIM, this portion of the scale may not be appropriate for proxy assessment.

Agreement between FONE-FIM scores provided by the patient vs. a significant other was stronger for motor items (ICC=0.79) than for cognitive items (ICC=0.61) (Tepper et al. 1996)

Agreement between patient and proxy FIM scores was reported to be ICC = 0.88 for the motor FIM and 0.38 for cognitive FIM --poorest agreements were noted for items that were most subjective and required opinion/judgement. In cognitive areas, proxies tended to score patients lower than the patients did themselves, while in activity areas, proxies tended to score patients higher than the patients themselves. When patients were grouped according to severity, it was noted that among patients with severe TBI, proxies rated patients as less disabled that the patients themselves, while for less severely injured patients, the opposite was true (Cusick et al. 2000).

*results from studies within the population of individuals with ABI/TBI appear in italics

**Advantages.** The Functional Independence Measure is a widely used, well-accepted, generic measure of burden of care used in inpatient rehabilitation settings. In clinical assessment, the greater number of items and wider choice of responses per item may yield more detailed information on an individual basis than assessments, such as the Barthel Index, with fewer items and response options (Hobart et al. 2001).

**Limitations.** The reliability of the FIM is dependent upon the individual conducting the assessment. Training and education in administration of the test is a pre-requisite for good levels of inter-rater reliability (Cavanagh et al. 2000). Length of time and amount of training required to arrive at a consensus score, as recommended by the developers of the FIM, may have significant implications for the practical application of the FIM in clinical practice.
The use of a single summed raw score may be misleading as it gives the appearance of a continuous scale. Steps between scores, however, are not equal in terms of level of difficulty and cannot provide more than ordinal level information (Linacre et al. 1994). Kidd et al. (1995) suggested that one use the summed scores as though on an interval level scale while the individual items remain ordinal.

Kidd et al. (1995) suggest that the inclusion of items related to communication and cognition as well as the ranking of 7 levels of severity for each item make the FIM more sensitive and inclusive. However, the contribution of the cognitive subscale to the scale as a whole is questionable. It has been shown to have less reliability and responsiveness than either the motor FIM or the total FIM (Ottenbacher et al. 1996; van der Putten et al. 1999; Hobart et al. 2001).

In an evaluation of responsiveness, FIM, motor FIM and the BI were all found to have similar effect sizes. The total-FIM was reported to exhibit no ceiling effect - 0% as compared to the BI’s 7% (van der Putten et al. 1999). This would suggest that the FIM might have no real advantage in terms of responsiveness to change despite having more items and a more precise scoring range for each item.

The FIM includes only 5 items to assess cognition. This limited cognitive assessment may be inadequate for the assessment of individuals who have experienced traumatic brain injury (Hall, 1994). In addition, the FIM is intended for use in assessed during inpatient rehabilitation and is not well suited to ongoing, long-term assessment in community-based settings (Gurkha et al. 1999).

Summary – Functional Independence Measure

Interpretability: The FIM has been well studied for its validity and reliability. It is widely used and has one scoring system increasing the opportunity for comparison. It is important to remember, when interpreting FIM scores, that it is an ordinal not continuous level scale.

Acceptability: Multiple modes of administration have been assessed including telephone interview. The FIM has been studied for use by proxy respondents.

Feasibility: Training and education of persons to administer the FIM, in addition to the price of the scale itself, may represent significant cost. Use of interview formats may make the FIM more feasible for longitudinal assessment.

---

**Table 17.20 FIM Evaluation Summary**

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
<th>Floor/ceiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
<td>Results</td>
</tr>
<tr>
<td>+++</td>
<td>+++ (TR)</td>
<td>+++ (IO)</td>
<td>+++ (IC)</td>
</tr>
</tbody>
</table>

*NOTE: +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver; varied (re. floor/ceiling effects; mixed results)*
17.9 Functional Assessment Measure (FIM+FAM)

The Functional Assessment Measure was created specifically for use with brain-injured patients in an attempt to enhance the appropriateness of the FIM for this specific population (Hall et al. 1993, Alcott et al. 1997, Hobart et al. 2001). The FIM contains only 5 cognitive items, which may limit its content validity in TBI populations (Hall, 1994). The FAM does not stand alone as an assessment tool, but rather consists of 12 items that are added to the 18 FIM items. The 12 additional items were developed by a team of clinicians representing each of the disciplines in a rehabilitation model (Hall et al. 1993) and are intended to emphasize cognitive, communicative and psychosocial function (McPherson et al. 1996).

The twelve FAM items are swallowing, car transfer, community access, reading, writing, speech intelligibility, emotional status, adjustment to limitations, employability, orientation, attention and safety judgement. Each item is rated using the same 7-point scale used on the FIM. Like the FIM, the FIM+FAM also consists of two subscales, one representing physical or motor functioning and one representing cognitive/psychosocial function. The total score for the FIM+FAM is 210, 112 for the motor FIM+FAM and 98 for the cognitive subscale (Gurka et al. 1999). Higher scores signify greater independence.

The FIM must be purchased from UDS and use of the FIM requires training and certification. The FAM items are in the public domain and can be downloaded from http://tbims.org/combi. A FIM+FAM rating form is available along with decision trees, training and testing vignettes specific to the FAM items from the website. The FIM+FAM requires approximately 35 minutes to administer (Hall, 1994).

<table>
<thead>
<tr>
<th>Table 17.21 Characteristics of the FIM+FAM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td><strong>Test-retest:</strong> ICC = 0.98 (Hobart et al. 2001)</td>
</tr>
<tr>
<td><strong>Interobserver reliability:</strong> Interrater (untrained raters) agreement reported to be 67% for FAM items and 55% for patients at admission to rehabilitation (Hall et al. 1993); agreement between raters was less than 70% for 7 items -- κ values for FIM+FAM items ranged from 0.35 – 0.95, FAM items ranged from 0.35 (adjustment to limits) to 0.92 (swallowing) – possible observer bias was identified for 4 items; employability, writing, comprehension and problem-solving (McPherson et al. 1996); at the item level, interrater reliability ranged from ICC= 0.36 (social interaction) to 0.97 (transfer – toilet, transfer – bed/ chair/ wheelchair &amp; stairs) – the average ICC for motor FIM+FAM = 0.91 and cognitive FIM+FAM = 0.74 – ICC values for total FIM+FAM = 0.83 (Donaghy &amp; Wass, 1998)</td>
</tr>
<tr>
<td><strong>Internal consistency:</strong> $\alpha = 0.96$ for FIM+FAM total, 0.96 for FIM+FAM motor and 0.91 for FIM+FAM cognitive/social items – item-to-total correlations ranged from 0.40 – 0.81 for FIM + FAM &amp; mean inter-item correlation = 0.46 (Hobart et al. 2001); $\alpha = 0.99$ for motor scale, 0.98 for cognitive scale and 0.99 for total FIM+FAM (Hawley et al. 1999)</td>
</tr>
<tr>
<td><strong>Validity</strong></td>
</tr>
<tr>
<td><strong>Construct validity:</strong> FIM+FAM not unidimensional -- factor analysis demonstrated 2</td>
</tr>
</tbody>
</table>
principal components with eigenvalues >1 – 16 items reflect physical functioning and 14 items reflect cognition, language and psychosocial functioning (Hawley et al. 1999); linear regression analysis revealed FIM+FAM cognition scores at 6 months explained 33% of variance in CIQ scores at 6 months post-discharge while FIM+FAM motor scores accounted for 22% of variance (this was compared to FIM cognition & motor scores that accounted for 31% & 21% of the variance, respectively) (Gurka et al. 1999)

Concurrent validity: Hall et al. (1993) reported FIM+FAM motor scores correlated with FIM motor (r=0.992) and FIM cognition scores (r=0.645) as well as with DRS ratings (r=0.680) – FIM+FAM cognition scores correlated with FIM motor (r=0.652), FIM cognition (r=0.952), DRS ratings (r=0.746), and LCFS scores (r=0.626); FIM+FAM correlated with BI (r=0.525, p<0.001), OPCS Index (r=0.824, p<0.001) and with the original FIM (r=0.962, p<0.001; McPherson & Pentland, 1997);

Concurrent validity (Convergent/Discriminant): FIM+FAM total and motor FIM+FAM scores correlated more strongly with OPCS disability scores, LHS scores, SF-36 physical component scores and WAIS-verbal IQ than with measures of mental health or psychological distress (SF36 mental component, General Health Questionnaire); however, cognitive FIM+FAM correlated most strongly with OPCS Disability scores and WAIS-verbal IQ scores and weakly with LHS, SF36 physical and mental components, and the General Health Questionnaire (Hobart et al. 2001)

Responsiveness
When ceiling effect is defined at scoring ≥180 on the FIM + FAM, 34% of patients scored in the ceiling range at discharge from rehabilitation and 79% at one year post discharge – this represented an improvement over the FIM (49%, 84%) – there was no advantage in terms of ceiling effect seen with regard to cog-FIM and the cognitive items of the FIM+FAM (Hall et al. 1996); Rasch analysis revealed FAM items cover a wider range of difficulty than the FIM items and, therefore, expand the range of scale difficulty beyond the FIM alone – however, both FIM and FAM items tend to cluster in the mid-range (Hall et al. 1993); 2% of patients were reported to obtain maximum score on FIM+FAM (McPherson & Pentland 1997); 80 – 90% of patients obtained “near maximum” scores on the FAM (Gurka et al. 1999) SRM means for FIM+FAM total, motor FIM+FAM and cognitive FIM+FAM were reported to be 0.42, 0.52 and 0.19 respectively – there were no significant floor or ceiling effects reported for FIM+FAM (Hobart et al. 2001)

Tested for ABI/TBI patients? Developed specifically for ABI/TBI population.

Other Formats
UK FIM+FAM – version of the FIM+FAM adapted for use in the United Kingdom resulting in a revised manual and decision trees for items identified as particularly difficult to score. Accuracy of scoring by individuals (when compared to a vignette with previously determined “correct” scores) was reported to be 77.1%. Accuracy of team scoring was reported to be 86.5% for the total score. Revision of the manual & decision trees increased accuracy of scoring for items perceived as difficult to score (Turner-Stokes et al. 1999).

Use by proxy? N/a

Advantages. The FIM was intended specifically for assessment during inpatient rehabilitation. The FAM items are better suited to evaluation post discharge from inpatient rehabilitation and may extend the applicability of the scale beyond the timeframe of the original FIM (Gurka et al. 1999). Addition of the FAM items to the FIM appeared to expand the range of abilities assessed (Hall et al. 1993).
Limitations. Use of the FIM+FAM still requires the use of trained raters who, ideally complete ratings after a period of observation and contribute to a team consensus process, as for the FIM (Hobart et al. 2001). The use of untrained raters may result in lower scale reliability (Hall et al. 1993).

Many of the FAM items have been identified as difficult to score (adjustment to limitations, emotion, employability, community mobility, safety judgement, attention and speech intelligibility (Turner-Stokes et al. 1999). Items in the expanded psychosocial/cognitive subscale seem to include more abstract concepts requiring raters to make more subjective assessments than was necessary for the more objective and observable behavioural items included on the original FIM (Hall et al. 1993, McPherson et al. 1996). Abstractness of items could have a deleterious effect on the reliability of those items (Alcott et al. 1997). Additional training together with more explicit definitions and/or content modification of the most abstract items could assist raters in the provision of reliable evaluations (Allcott et al. 1997, McPherson et al. 1996).

While the FAM items were intended to provide additional assessment of the psychosocial aspects of disability following brain injury (Hall et al. 1993), the validity of the assessment has not been clearly established (Hobart et al. 2001). The psychosocial/cognitive FIM+FAM does not correlate well with measures of handicap, such as the London Handicap Scale or as strongly as one might expect with the mental component summary of the SF36 (Hobart et al. 2001). Overall, the added length and increased training requirements associated with the FIM+FAM do not seem to offer any substantial advantage over the FIM (McPherson & Pentland 1997, Hobart et al. 2001). While the FIM+FAM appears to evaluate a somewhat broader range of abilities (Hall et al. 1993), reports of ceiling effects associated with the FIM+FAM are varied and reported effect sizes are approximately the same as those reported for the FIM (Hobart et al. 2001).

Summary –FIM+FAM

Interpretability: The 18-FIM items are widely used and recognized. However, the FAM items are more difficult to rate reliably and the validity of FAM is not well established.

Acceptability: Alternate modes of administration have not been examined and FAM items have not been evaluated for use in assessment by proxy.

Feasibility: The addition of FAM items to the FIM creates a longer assessment requiring the involvement of addition raters in team consensus and more training for these raters. While the FAM items are freely available use of the FIM items requires purchase of the scale, training and certification.
Table 17.22 FIM+FAM Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>++</td>
<td>+++ (TR)</td>
<td>++</td>
</tr>
<tr>
<td>++(IO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+++(IC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.

17.10 Galveston Orientation and Amnesia Test (GOAT)

The Galveston Orientation and Amnesia test (GOAT) was intended to evaluate orientation to time, place and person and to provide an estimation of the intervals prior to and following the injury for which there is no recall (Levin et al. 1979).

Assessment consists of 10 items regarding orientation to person (name, address & birthdate), place (city/town and building they are in) and time (current time, date, month, year & date of hospital admission) as well as memory of events both after and prior to the injury (Bode et al. 2000). Oral questions are posed directly to the patient who may respond either orally or in writing (Levin et al. 1979, Jain et al. 2000). Error points are awarded for each incorrect response, summed and deducted from 100 to arrive at the total score. Both the scale and instructions for assigning error points are available in Levin et al. (1979).

The duration of PTA is defined as the period following coma in which the GOAT score is less than 75 (Levin et al. 1979). PTA is considered to have ended if a score of 75 or more is achieved on 3 consecutive administrations (Wade 1992, Novack et al. 2000, Zafonte et al. 1997).

In the initial standardization study of Levin et al. (1979) using patients with mild head injury as a reference group, it was determined that a score of 75 represented a level achieved by 92% of the standardization group. No patients with mild head injury scored less than 65 on the GOAT. Scores between 66 and 75 are considered borderline-abnormal while scores above 75 fall into the range considered normal within the reference group (Levin et al. 1979; van Baalen et al. 2003).

The GOAT is a brief and simple mental status examination developed for use by health professionals at the bedside or in the Emergency Department (Levin et al. 1979; van Baalen et al. 2003).

Table 17.23 Characteristics of the Galveston Orientation and Amnesia Test

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Interobserver reliability: Kendall tau = 0.99 (p&lt;0.001) and 0.99 for individual items (Levin et al. 1979)</th>
</tr>
</thead>
</table>

Internal consistency: on Rasch analysis, person separation reliability = 2.46 and item separation reliability = 8.68 and all items adhered to a single construct (Bode
Validity

Construct validity: on Rasch analysis, the constructed item hierarchy confirmed previous research – orientation to person, place & time precedes aspects dealing with memories surrounding the injury (Bode et al. 2000); GOAT scores correlated positively with GCS scores (r=0.456, p<0.002) and with admission and discharge FIM scores (r=0.701 and 0.531, respectively; Novack et al. 2000); Levin et al. (1979) reported impaired eye opening on the GCS strongly related to regaining orientation as measured on the GOAT(χ² = 21.09; p<0.00001), GCS motor responding and subsequent GOAT performance (χ² = 18.98; p<0.00001) and GCS verbal responding and persistence of amnesia as assessed by GOAT (χ² = 19.53; P<0.00001). Levin et al. (1979) also demonstrated GOAT performance associated with CT findings (p=0.02);

Concurrent validity: Scores on GOAT and JFC PTA scale reported to be strongly correlated (r=0.99; p<0.000) (Forrester et al., 1994); GOAT scores correlated with Orientation Log scores (r=0.901, p<0.001) (Novack et al. 2000); GOAT assessment of PTA correlated with Length of coma (r² = 0.575, p<0.0001), however, this relationship varied with severity of injury and duration of coma (Katz and Alexander, 1994)

Predictive validity: PTA, measured by GOAT scores, significant independent predictor of functional outcome (p=0.00005) as assessed by DRS and FIM total, motor and cognitive scores (Zafonte et al. 1997); length of PTA as assessed by GOAT significantly associated with employment status at 1 year post-injury (Cifu et al. 1997); Levin et al. (1979) reported GOAT performance association with long term outcome (at least 6 months post-injury) rated on the GOS (p<0.0001); Katz & Alexander (1994) reported GOAT scores (PTA) to be associated with GOS at 6 and 12 months post injury (r² = 0.447 and 0.476, respectively, p<0.0001) among patients with diffuse axonal injury.

Responsiveness

N/A

Tested for ABI/TBI patients?

Designed for use with TBI patients

Other Formats

A-GOAT - developed specifically for use with aphasic patients (Jain et al. 2000) – essentially, the GOAT in a multiple choice format – allow for comparison of aphasic and nonaphasic patients on the same standard – 10 items with a 3-choice response format

- using GOAT as the gold standard, the AGOAT (cut-off >90) demonstrated 100% sensitivity and 95% specificity in identifying PTA

Use by proxy?

N/A

Advantages. The GOAT provides an objective rating of early cognitive recovery eliminating the need for sometimes ambiguous terminology used to describe mental status, such as “confused” (Levin et al. 1979). Rasch analysis demonstrated that items on the GOAT represents a wide range of difficulty suggesting that the scale is useful for assessing patients with a wide range of cognitive impairments (Bode et al. 2000).

Limitations. The standard GOAT response format makes administration difficult with nonverbal patients (Novack et al. 2000). The requirement for oral or written expression may result in penalizing patients who are experiencing deficits of expression but not in
orientation or in the retrieval or consolidation of memory (Jain et al. 2000). An aphasia-specific version of the GOAT has been created, however, it requires further evaluation.

For items in which partial credit is used, Rasch analysis revealed step disorder (Bode et al. 2000). Collapsing these response categories to a simple dichotomy (right vs. wrong) eliminated the disorder and allowed the construction of an equal interval measure from the GOAT (Bode et al. 2000).

While the GOAT does contain items intended to provide an assessment of memory, it is primarily a measure of disorientation. Eight of the 10 GOAT items evaluate orientation while only two examine memory (Forrester et al. 1994).

**Summary – GOAT**

**Interpretability:** The GOAT provides an objective assessment with a standardized cut-off for the presence of PTA.

**Acceptability:** In its original form, the GOAT is not well suited to assessment of patients with aphasia.

**Feasibility:** The GOAT may be too lengthy for a simple, repeated bedside assessment of mental status. However, it is freely available and can be used by any healthcare professional.

### Table 17.24 GOAT Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
<th>Floor/ceiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor Results</td>
<td>Rigor Results</td>
</tr>
<tr>
<td>+</td>
<td>+++(IO)</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>+</td>
<td>++ (IC)</td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC=internal consistency; IO=Interobserver;
The Glasgow Coma Scale (GCS) was developed as a simple, objective assessment of impaired consciousness and coma based on eye opening, verbal and motor responsiveness (Teasdale & Jennett 1974, Teasdale & Jennett 1976, Teasdale et al. 1978). It has become the most widely known and widely used scale in the assessment of level of consciousness (Wade 1992, Hall 1997, Brain trauma foundation, 2000).

The GCS is an observer rating scale consisting of 15 items in three basic categories: motor response (6 items), verbal response (5 items) and eye opening (4 items). Points are awarded for the best response in each category and category scores are summed to provide a global GCS score (Wade1992, Sternbach 2000).

Additional categorical divisions are used to differentiate patients in terms of initial severity of head injury such that GCS scores of 13–15 represent mild injury, scores of 9–12 represent moderate injury and scores of $$\leq8$$ represent severe injury (Sternbach 2000). The GCS is freely available, takes approximately 1 minute to administer and can be performed by all medical personnel (Oppenheim & Camins 1992).

**Table 17.25 Glasgow Coma Scale (Teasdale and Jennett, 1974)**

<table>
<thead>
<tr>
<th>Response/Item</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eye Opening:</strong></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>4</td>
</tr>
<tr>
<td>To speech</td>
<td>3</td>
</tr>
<tr>
<td>To pain</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td><strong>Motor Response:</strong></td>
<td></td>
</tr>
<tr>
<td>Obey commands</td>
<td>6</td>
</tr>
<tr>
<td>Localizes pain</td>
<td>5</td>
</tr>
<tr>
<td>Withdrawal (from painful stimulus)</td>
<td>4</td>
</tr>
<tr>
<td>Abnormal flexion</td>
<td>2</td>
</tr>
<tr>
<td>Extension</td>
<td>1</td>
</tr>
<tr>
<td>None</td>
<td></td>
</tr>
<tr>
<td><strong>Verbal Response:</strong></td>
<td></td>
</tr>
<tr>
<td>Oriented</td>
<td>5</td>
</tr>
<tr>
<td>Confused</td>
<td>4</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>3</td>
</tr>
<tr>
<td>Incomprehensible</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>1</td>
</tr>
</tbody>
</table>

Summed scores range from a minimum of 3 (total un-responsiveness) to a maximum of 15 (alert, fully responsive). A total of 8 or less is used to separate coma from non-coma (Wade 1992).

**Table 17.26 Characteristics of the Glasgow Coma Scale**

| Reliability | Interobserver reliability: $\kappa$ ranged from 0.39 – 0.79 (p=0.000; Juarez & Lyons, 1995); overall agreement = 90% -- for each component, % agreement ranged from 83.8% (eye opening-right) to 98.7% (best motor response-left) – agreement was lowest for the eye opening component; correlations between observers ranged from 0.855 (motor response-right) to 0.974 (best verbal response) (Fielding & Rowley, 1990); low rates of disagreement (disagreement rating = 0.089 – 0.187), motor responses elicited by supraorbital stimulus had higher rates of disagreement than fingertip stimulation (p<0.01; Teasdale et al. 1978); % agreement ranged from 55% (verbal) - 74% (eye opening), Spearman’s rho = 0.587 (verbal) – 0.742 (motor) and $\kappa_w = 0.48$ (verbal) to 0.72 (eye opening) (Gill et al. 2004); Rowley & Fielding (1991) reported reliability of 98.6 – 100% among experienced nurses, 94.3% - 96.2% among new graduates and 77.3% - 100% among groups of student nurses. |
Validity

**Construct validity:** GCS 13 – 14 associated with greater proportion of abnormality on CT and longer duration PTA than GCS 15 (McCullagh et al 2001); Depth of coma assessed by GCS is considered to reflect extent of brain damage, in 1996 review, Prasad cited the following studies -- good correlation between GCS and cerebral metabolic rate (Langfitt et al. 1982), correlation with CSF enzymes (0.82 – 0.99; Bakay et al., 1983) and good correlation with evoked potential abnormalities (no stat given; Lindsay et al. 1981); mean values of serum enzymes LDH, and CPK correlated with GCS scores within 72 hours of injury (r=0.89 for both), incidence of multiple trauma also correlated with GCS scores (p<0.01; Bakay et al. 1983);

**Concurrent validity:** GCS scores correlated with length of coma (r = 0.233, p<0.0001; Katz and Alexander, 1994)

**Predictive validity:** GCS 13-14 and GCS 15 (mild head injury) not predictive of neuropsychiatric outcome 6 months post-injury (McCullagh et al. 2001); On multiple regression, GCS was identified as a significant independent predictor of death (p<0.0001), however, the prognostic value of the GCS was noted to be affected by mechanism of injury and age of the patient (Demetriades et al. 2004); based on 10 years of head injury data (1992 – 2001), the GCS was significantly correlated with the GOS at 6 months post injury for each year of data from 1992 – 1996, but from 1997 – 2001, no significant association was reported (Balestreri et al. 2004); GCS was predictive of survival (AUC=0.891) but only slightly more than the motor component score on its own (AUC=0.873), the eye opening score did not add to the predictive accuracy of the GCS (Healey et al., 2003); In predicting mortality, there was a significant association between total GCS scores and outcome – on multivariate analysis, the motor & verbal components were associated with mortality while eye-opening was not – additionally, among patients with total GCS > 9, only the verbal component was significant on multivariate analysis, whereas for patients with GCS≤9, Motor and Verbal component scores were significantly associated with mortality and verbal score was a better predictor than motor score in this group (Teoh et al. 2000); initial GCS scores significantly associated with employment status at one year post-injury(p<0.05)(Cifu et al. 1997); initial GCS significantly associated with DRS scores, LCFS scores, FIM-motor and FIM-cognitive at admission to and discharge from rehabilitation, though correlations were low to moderate (r = 0.16 to 0.37; all p<0.0005; Zafonte et al. 1996); Waxman et al. (1991)reported that, when taken immediately on arrival at hospital, reported correlations between GCS scores and GOS scores (r² =0.16) as well as length of hospital stay (r² = 0.08), length of intensive care stay(r² = 0.05) and duration of ventilatory support (r² =0.03) were low, however, correlations between GCS taken at 6 hours after hospital arrival and GOS scores were much stronger (r² = 0.55) – GCS assessed at 6 hours and change in GCS contributed significantly to the prediction of GOS (r² = 0.71 – model included GCS 6 hours, Initial severity score, number of abnormal CT findings & change in GCS score); GCS was predictive of GOS at 6 months (r² = 0.135, p<0.001) but much less so at 12 months (r² = 0.08, p<0.005; Katz and Alexander, 1994); 95% of patients scoring higher than 7 on initial GCS had favourable GOS outcome, 95% with GCS lower than 5 had unfavorable outcome – prediction of outcome for patients with initial GCS of 5,6,7 was more difficult – 24 hours GCS scores were preferable among these middle band patients when patients had either improved or deteriorated into the range in which predictions were more accurate (Young et al. 1981); best day 1 GCS was a significant predictor of 6 month outcome on the DRS (p<0.05; Pastorek et al. 2004); initial Glasgow Outcome Score reported as a significant predictor of GOS score 6 months post injury (p<0.001; Satz et al. 1998); preresuscitation GCS score correlated with survival for head injury patients


**Advantages.** The Glasgow Coma Scale is a simple, straightforward and very brief bedside assessment. It is the most widely used instrument in the assessment of level of consciousness. GCS scores are a significant predictor of outcome following head injury; however, the prognostic value of the GCS is increased by taking other variables into account as well, such as mechanism of injury, age, CT findings, papillary abnormalities and episodes of hypoxia and hypotension (Demetriades et al. 2004, Balestrieri et al. 2004, Zafonte et al. 1996).

**Limitations.** The GCS is based on the assumption that evaluation of eye opening sufficient to represent brainstem arousal systems activity. While other assessments have been developed to provide a more comprehensive evaluation of brainstem responses, the resulting tools are substantially more complex than the GCS (Sternbach 2000).

The GCS has been reported to be reliable when used by various groups of healthcare professionals regardless of level of education or ICU experience (Juarez and Lyons 1995). Nurses and general surgeons have been reported to be as consistent in their ratings as neurosurgeons (Teasdale et al. 1987). However, it has also been demonstrated that consistent ratings may also be inaccurate. Rowley and Fielding (1991) reported that the percentage agreement between inexperienced individuals and expert raters ranged from 58.3% to 83.3%. Lower levels of accuracy were most notable in the middle ranges of the scale. Training and the implementation of standard assessment procedures are important to maintain both high levels of reliability and accuracy of evaluation. The administration of a painful stimulus appears to be somewhat controversial and there is a great deal of variability in the means and location of its application (Edwards 2001, Lowry 1999).

The GCS is most often reported as a single, overall score although the scale authors did not recommend the summary score for use in clinical practice. While the single, global score may be a convenient way to summarize data, the use of a global score may result in a loss of information that adversely affects the predictive accuracy of the GCS (Teasdale et al. 1983, Teoh et al. 2000, Healey et al. 2003). The use of a global summary score assumes that each category is equally weighted (Teasdale et al. 1983). However, it has been reported that motor response has the greatest influence on the summary score and results are skewed toward this component (Bhatty and Kapoor 1993). Healey et al. (2003) demonstrated that the ability of the GCS score to predict survival was

---

### Table: Glasgow Coma Scale Evaluation

<table>
<thead>
<tr>
<th>Responsiveness</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested for ABI/TBI patients?</td>
<td>Created specifically to monitor head-injured patients</td>
</tr>
<tr>
<td>Other Formats</td>
<td>N/A</td>
</tr>
<tr>
<td>Use by proxy?</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(r=0.978; p<0.0001) and with functional outcome as assessed by the FIM at discharge (r=0.973, p<0.0001; (Udekwu et al. 2004).
derived mostly from the motor response category. In addition, the summary score represents a potential 120 combinations of scores from the three GCS components collapsed into only 13 possibilities. Different combinations of motor responsiveness, verbal responsiveness and eye-opening may have different associated outcomes. Teoh et al. (2000) reported significant differences in mortality outcomes between 4 of 11 scores with multiple permutations demonstrating that individuals with the same GCS scores in varying permutations can have significantly different risks for mortality.

Perhaps the most frequently encountered limitation of the GCS is untestable components in various patient groups. Pastorek et al. (2004) reported that the ability of the patient to be evaluated on the entire GCS contributed to the prediction of global outcome measures at 6 months (Pastorek et al. 2004) Unfortunately, patients who have been intubated or sedated, those with paralysis or facial swelling, patients with hypotension, hypoxia, alcohol or illicit drug intoxication all may not be able to provide responses to all categories of GCS items for reasons unrelated to head trauma (Oppenheim and Camins 1992, Rutledge et al. 1996, Demetriades et al. 2004). Murray et al. (1999 as cited in Teasdale and Murray 2000) reported that in a study of head injury patients in European centres, total assessment was possible in 61% of patients before hospital, in 77% on arrival at hospital and in 56% of patients arriving at a neurosurgical unit. It has been suggested that inability to assess using the GCS may reflect the increased and more aggressive use of intubation, ventilation and sedation (Teasdale and Murray 2000, Balestreri et al. 2004). When the GCS was developed, the initial assessment was to be undertaken approximately 6 hours after injury to allow time for stabilization of systemic problems, but prior to the initiation of interventions such as neuromuscular paralyzing agents or sedatives (Bakay et al. 1983, Marion and Carlier, 1994). Increasingly, GCS assessment is performed upon arrival at the Emergency Department and many patients are already intubated and/or sedated by that time (Marion and Carlier, 1994, Waxman et al. 1991).

Summary – GCS

**Interpretability:** The GCS is the most familiar, most widely-used early assessment of level of consciousness. It has established categories related to the presence of coma and severity of injury.

**Acceptability:** A very brief, simple observer rater scale. The application of painful stimulus is controversial. Assessment of all components is compromised by aggressive, early interventions such as intubation and sedation.

**Feasibility:** The scale is simple to administer and designed for use by any health profession. Lack of experience and variability in assessment may result in inaccurate assessment. Training and standardized procedures are recommended.
Table 17.27 GCS Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor Results</td>
</tr>
<tr>
<td>+</td>
<td>++ (IO)</td>
<td>++</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.

17.12 Glasgow Outcome Scale (GOS)/Extended Glasgow Coma Scale (GOSE)

The Glasgow Outcome Scale (see below) is a practical index of social outcome following head injury designed to complement the Glasgow Coma Scale as the basis of a predictive system (Jennett and Bond, 1975, Jennett et al. 1981). It is a simple, hierarchical rating scale with a limited number of broad categories. The scale focuses on how head injury had affected function in major life areas and is not intended to provide detailed information on specific deficits (Wilson et al. 1998). Individuals within any single outcome category represent a range of abilities (Jennett and Bond, 1975).

Patients are assigned to one of 5 possible outcome categories: death, persistent vegetative state, severe disability, moderate disability, and good recovery (Jennett and Bond, 1975). The GOS is described in the following table. In 1981, a revision to the scale was proposed to better classify patients who had regained consciousness (Jennett et al. 1981). In the Extended Glasgow Outcome Scale (GOSE), each of the 3 categories applicable to conscious patients are subdivided into an upper and lower band resulting in 8 possible categories. GOS ratings can be derived from the GOSE by collapsing these subdivisions (Wilson et al. 2000).

The assignment of an individual to an outcome category should be based on the results of a structured interview focused on social and personal functional ability (Jennett et al. 1981). The final rating is based on the lowest category of outcome indication in the interview (Wilson et al. 2000).

### Table 17.28 Glasgow Outcome Scale

<table>
<thead>
<tr>
<th>Category</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dead</td>
<td></td>
</tr>
<tr>
<td>Persistent Vegetative State</td>
<td>Sleep/awake, nonsentient</td>
</tr>
<tr>
<td>Severely Disabled*</td>
<td>Conscious but dependent</td>
</tr>
<tr>
<td>Moderately Disabled*</td>
<td>Independent but disabled</td>
</tr>
<tr>
<td>Good Recovery*</td>
<td>May have mild residual effects</td>
</tr>
</tbody>
</table>

*Teasdale et al. 1998. **Note: The categories of severely disabled, moderately disabled and good recovery were subdivided into upper and lower bands to form the Extended GOS (Jennett et al. 1981)**

### Table 17.29 Characteristics of the GOS and GOSE

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest: $\kappa$ ranged from 0.40 – 0.92 for the GOS and 0.40 – 0.87 for the GOSE; however, the retest period was lengthy ranging from 3 – 6 months (Maas et al. 1983)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Interobserver reliability:</strong> Jennett et al. (1981) reported 95% agreement between observers using the original GOS; agreement between assessment based on a mail-administered research questionnaire and assessment via interview by a</td>
</tr>
</tbody>
</table>

---

**NOTE:** The categories of severely disabled, moderately disabled and good recovery were subdivided into upper and lower bands to form the Extended GOS (Jennett et al. 1981)**
psychologist was reported to be $r=0.79$ while agreement between a GP’s assessment and the psychologist interview was $r=0.49$ (Anderson et al. 1993); based on live interviews $\kappa=0.77$ for GOS and 0.48 for GOSE – when ratings were based on previously recorded data, $\kappa=0.58$ for GOS and 0.49 for GOSE – agreement between live and recorded data ratings, $\kappa=0.77$ for GOS and 0.53 for the GOSE(Maas et al. 1983); 70% of GOS ratings were in perfect agreement while none differed by more than one category – for the GOSE none differed by more than one category –most discrepancy seen in the middle categories (Brooks et al. 1986)

### Validity

**Construct validity:** GOS ratings reported to be associated with results neurological testing of motor tasks ($p<0.001$), psychomotor tests ($p<0.05$), assessments of memory variables ($p<0.05$), attention variables ($p<0.05$) such that neuropsychological test performance decreased as a function of increased severity on the GOS rating scale (Satz et al. 1998); performance on cognitive tests 3 months post injury differed significantly ($p<0.05$) between outcome subgroups corresponding to GOS ratings demonstrating a clear gradation in cognitive scoring between groups in the expected direction – this relationship was not as clear when the GOSE was used (Brooks et al., 1986)

**Construct validity (known groups):** GOS scores could discriminate between groups based on categories of vocational recommendations (return to work, vocational training, supported work and continued remedial therapy; $p<0.0001$), GOS scores accounted for 76% variance between cell means (Mysiw et al. 1989)

**Concurrent validity:** Admission DRS scores correlated with initial Stover& Zeiger (S-Z) ratings ($r=0.92$), discharge DRS scores correlated with discharge SZ scores ($r=0.81$), GOS scores (0.80) and EGOS scores (0.85) (Gouvier et al. 1987); GOS ratings correlated with SF-36 subscale scores ($r=0.51 – 0.68$, $p<0.01$; Jenkinson et al. 1993 cited in Teasdale et al. 1998); GOS scores correlated with DRS ratings at admission to ($r=0.50$, $p<0.01$) and discharge from rehabilitation ($r=0.67$, $p<0.01$; Hall et al. 1985)

**Predictive validity:** GOS at discharge from rehabilitation significantly correlated with GOS 5 – 7 years after head injury ($r=0.60$, $p<0.001$) and with discharge destination ($p<0.0001$; Massagli et al. 1996)

### Responsiveness

From assessment 3 months post injury to 6 month assessment 36% of patients demonstrated change in GOSE ratings while only 11% demonstrated change in category based on GOS ratings ($p<0.05$) (Levin et al. 2001); from admission to discharge from rehabilitation, improvement shown by the DRS was significantly greater than that shown by the GOS (71% vs. 33%, $p<0.01$; Hall et al. 1985)

### Tested for ABI/TBI patients?

Specific to head injury populations.

### Other Formats

**Structured interview for the GOS/GOSE (Wilson et al. 1998)** – improves reliability and removes limitations associated with scale ambiguity and lack of guidelines for administration. – specifies criteria for separating the upper and lower bands of the upper 3 categories of the GOS. The structured interview consists of a series of questions regarding consciousness, independence (both at and away from home), social roles (work, social activities, leisure, relationships) and return to normal life(Wilson et al. 2000). The questionnaire focuses on aspects of social disability (effects on social & leisure activities and disruption to family and friendships) as originally described by Jennett et al. (1981). The structured interview format also allows for the inclusion of pre-injury disability status (Wilson et al. 1998) and provides specific guidance regarding assignment to outcome category (Teasdale et al. 1998).
Reliability: agreement between raters was reported to be 92% for the GOS and 78% for the GOSE when administered via structured interview, $\kappa_w = 0.89$ and 0.85 for the GOS & GOSE respectively. (Wilson et al. 1998)

Validity: significant correlation reported between BI and GOS ($\rho=0.61$, $p<0.001$) and between DRS scores and GOS ratings ($\rho=0.89$, $p<0.001$; Pettigrew et al. 1998); when using the structured interview, Wilson et al. (2000) reported correlations with BI scores of 0.47 & 0.46 for the GOS and GOSE, respectively – GOS and GOSE ratings also correlated with DRS ratings ($r=0.89$ for both), in addition, GOS & GOSE ratings correlated with Beck Depression inventory scores ($r=0.61 & 0.64$), GHQ scores (0.57 & 0.59), SF-36 subscores (ranging from 0.41 – 0.67 and 0.47 – 0.71) and NFI scale scores, ranging from 0.33 – 0.57 and 0.37 – 0.63 for patient NFI ratings and 0.47 – 0.68 and 0.47 – 0.69 for NFI ratings obtained from friends or relatives; Levin et al. (2001) reported that at 3 months post injury, GOS ratings were significantly associated ($p \leq 0.05$) with results on the CES-D, CIQ, Social support questionnaire, and the paced auditory serial audition test (trial 1) – GOSE ratings were significantly associated with results from the CIQ & the Paced auditory serial audition test (trial 1) – in cases where both demonstrated linear association with scale scores (i.e. CIQ and the paced auditory serial audition test) GOSE ratings accounted for more of the variance in scale scores than GOS ratings ($r^2 = 0.35$ vs. 0.26 and 0.37 vs. 0.19, respectively).

Telephone administration (structured interview): agreement between face-to-face-interview and telephone interview was reported to be $\kappa_w = 0.92$ for the GOSE – when GOSE scores were collapsed to GOS ratings, $\kappa_w = 0.92$ – interobserver agreement was reported to be $\kappa_w = 0.84$ and 0.85 when GOSE scores were collapsed to GOS ratings (Pettigrew et al. 2003)

Simple Postal Assessment (Hellawell et al. 2000): using a simple, 4 question survey, inter-observer (GP’s, family informants, experienced GOS raters) reliability was reported to range from $\kappa = 0.17$ (between GP and experienced rater) to 0.61 (between GP’s and family informants).

Postal Questionnaires – based on the structured interviews for GOS and GOSE of Wilson et al. (1998) [Wilson et al. 2002]: designed to be completed by the patient or a relative or caregiver of the patient or by the patient with the assistance of a significant other/caregiver. Questions are intended to discriminate between the categories of severe disability, moderate disability and good recovery (for the GOSE questionnaire, these are further subdivided into upper and lower bands). Return rates were reported to be 76% for the GOS questionnaire and 83% for the GOSE questionnaire. Test retest reliability for the GOS was reported to be $\kappa_w = 0.94$ and $\kappa_w = 0.98$ for the GOSE. Agreement between GOS ratings assigned via postal questionnaire and telephone interview (using the structured interview) was reported as $\kappa_w = 0.67$ while agreement using the GOSE questionnaire was higher ($\kappa_w = 0.92$)

Edinburgh Extended Glasgow Outcome Scale (EEGOS) (Hellawell & Signorini, 1997): based on the GOS, but requires scoring for behavioural/emotional, cognitive, and physical functioning – each patient is assigned a rating on each of these types of function. Descriptions are provided for each of the function types. Using retrospective data, interobserver agreement was reported as $\kappa = 0.20 – 0.55$ for behavioural ratings, $\kappa = 0.56 – 0.63$ for cognitive ratings, and $0.57 – 0.75$ for physical ratings. Using current data, interobserver agreement for behavioural, cognitive and physical ratings was reported as $\kappa=0.61$, 0.62 and 0.73, respectively (Hellawell & Signorini, 1997)

Use by proxy? recommended that the best source of information be used – whenever
possible include information gained by interviewing close friend or family member (Wilson et al. 1998); using a simple postal survey, GP’s and informants tended to rate patient outcome more positively than experienced GOS raters (Hellawell et al. 2000).

**Advantages.** The GOS is the most widely used and accepted measure of outcome following head injury (Wade 1992). It has been adopted widely for use in clinical trials (Wade 1992, Wilson et al. 2000, Hellawell et al. 2000). It is a simple, reliable means of describing recovery (Jennett et al. 1981) that is quick to administer, broadly applicable and has clinically relevant categories (Wilson et al. 2000).

Structured interviews and guidelines for their administration are available for the GOS and GOSE (Wilson et al. 1998. Each interview incorporates a means for including information regarding pre-injury status, thereby providing a means for determining the effect of the sequelae of head injury on outcome separate from the effects of pre-existing conditions or circumstances (Wilson et al. 1998, Pettigrew et al. 1998). While use of the structured interview has increased the reliability of postal and telephone administration, face-to-face interview remain the preferred method to determine a GOS rating (Wilson et al. 2002).

**Limitations** The GOS provides an overall assessment of outcome and does not provide detailed information with regard to specific disabilities or handicaps. Categories are broad and the scale does not reflect subtle improvements in functional status of an individual (Pettigrew et al. 1998). Individuals may achieve considerable improvement in ability, but not change outcome category (Brooks et al. 1986). The GOS rating was intended primarily to provide an overall summary of outcome and facilitate comparison not to describe specific areas of dysfunction (Pettigrew et al. 1998). In addition, GOS outcome categories are often expressed as a dichotomy; poor or unfavourable outcome versus independence or favourable outcome. This results in a loss of information and decreased sensitivity (Teasdale et al. 1998).

Originally, GOS categories were described according to a range of features, but specific criteria were not defined for each of the different outcomes. This lack of clarity may have had a negative impact on scale reliability by introducing an element of subjectivity on the part of the rater (Teasdale et al. 1998, Maas et al. 1983). In addition, attempts to increase the sensitivity of the GOS by subdividing the upper 3 categories in an upper and lower band was associated with decreased consistency in category assignments (Maas et al. 1983). However, the structured interview and guidelines created by Wilson et al. (1998) have alleviated much of the difficulty surrounding ambiguous assignment criteria.
Summary – GOS

**Interpretability:** The GOS is widely used and accepted. The GOS provides an overall assessment suitable for the comparison of outcomes at the group level.

**Acceptability:** The brevity and simplicity of the GOS facilitate patient compliance. The GOS has been studied for use by telephone and mail administration. Structured interviews improve the reliability of administration by these methods.

**Feasibility:** The GOS can be used by professionals from various backgrounds and does not require any physical, psychiatric or neurologic examination. It is well-suited to busy clinical settings and large scale research trials.

<table>
<thead>
<tr>
<th>Table 17.30 GOS/GOSE Evaluation Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Rigor</td>
</tr>
<tr>
<td>+</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.

17.13 Hospital Anxiety and Depression Scale (HADS)

The Hospital Anxiety and Depression Scale (HADS), a self-assessment scale, was developed to detect states of depression, anxiety and emotional distress amongst patients who were being treated for a variety of clinical problems (Zigmond and Snaith 1983). The scale was not designed to be a clinically diagnostic tool (Whelan-Goodinson et al., 2009). Originally the scale consisted of 8 questions relating to depression and 8 relating to anxiety. Initial findings indicated one of the items on the depression scale was weak (r=.11) thus it was removed. Remaining items on the scale had correlations ranging from +.60 to +.30, with a significance of (p<0.02). Anxiety items had correlations ranging from +.76 to +.41 (p<0.01); however, to keep the items in each scale equal, the weakest item on the anxiety portion of the scale was removed. Thus the final scale has a total of 14 items, with responses being scored on a scale of 0-3, with 3 indicating higher symptom frequencies (Whelan-Goodinson et al., 2009). Score for each subscale (anxiety and depression) can range from 0-21 with scores categorized as follows: normal (0-7), mild (8-10), moderate (11-14), severe (15-21). Scores for the entire scale (emotional distress) range from 0-42, with higher scores indicating more distress. Prior to completing the scale patients are asked to “fill it complete in order to reflect how they have been feeling during the past week” (Zigmond and Snaith, 1983; p. 366).

Unfortunately within the TBI population, many measures are used to assess depression and anxiety post injury, but none of these measures have been evaluated for use with this population (Whelan-Goodinson et al., 2009; Schonberger and Ponsford, 2010). Recently the HADS has been tested with those who have sustained an ABI; however due to the mixed aetiology problems were found with some of the questions which could be related to the injury itself, the level of cognitive impairment or the
decreased speed at which information is processed (Dawkins et al., 2006; Johnston et al., 2000).

Table 17.32 Characteristics of the HADS

| Reliability | Test-retest: Results indicate there is good test-retest reliability on the HADS at 0-2 weeks \( r = 0.84 \); >2-6 weeks \( r = 0.73 \); and >6 weeks \( r = 0.70 \) for the anxiety subscale. Results from the depression subscale were 0-2 weeks \( r = 0.85 \); >2-6 weeks \( r = 0.76 \); and >6 weeks \( r = 0.70 \); indicating the HADS was stable enough to withstand situation influences (Herrmann 1997). Pearson product movement correlation was found to be .92 and .90 between the HADS total score and the HADS anxiety score and the HADS depression score (Herrero et al., 2003). Inter-Rater Reliability: Kappa scores indicated there was no significant difference between the GHQ-28 and HADS (total score) (kappa statistic = 0.74, SE = 0.09, \( p = 0.04 \)). Internal consistency: Good internal consistency was found (\( \alpha = 0.80 \) for the anxiety subscale scale and \( \alpha = 0.81 \) for the depression subscale) during initial testing (Zigmond and Snaith, 1983). Whelan-Goodinson et al. (2009) found good internal consistency with Cronbach \( \alpha \) scores of .94 on the total HADS score, .88 on the depression subscale and .92 on the anxiety subscale. On other studies the internal consistency has been found to range from .68 to .93, mean .83 (for the anxiety subscale) and .67 to .90, mean .82 (for the depression subscale) (Bjelland et al., 2002). In an earlier study, Lisspers et al. 1997, found Cronbach \( \alpha \) scores for the HADS total score to be .84, for the HADS anxiety subscale, .82 and for the HADS depression subscale .90. Scores in this study were not affected by gender or age. Herrero et al. (2003) found, while validating the scale with a group of Spanish patients, Cronbach \( \alpha \) scores to be .90 for the full scale, .84 for the depression subscale and .85 for the anxiety subscale. Subscales also correlated with each other \( r = 0.68, P < 0.01 \) and each subscale correlated with the full scale \( r = 0.02, P < 0.01 \). |

| Validity | Convergent Validity: the correlation between the HADS depression subscale and the Beck Depression Inventory Primary Care (BDI-PC) has been found to be .62, \( p < 0.001 \) (Beck et al., 1997). Concurrent Validity: Higher scores on the HADS-depression subscale were linked to higher scores on the SCID-IV (mean=3.52; sd=3.01 and mean=9.29; sd 5.19 respectively; \( t = 6.84, df = 98, p < 0.001 \). Of note 38.2% diagnosed as depressed on the SCID-IV scored within the normal range on the HADS-D. Results from the SCID-IV for those diagnosed with an anxiety disorder (mean 11.42, sd 4.75) had a higher mean score on the HADS anxiety subscale (mean 5.37, sd 3.95) \( (t = 6.47, df = 62.41, p = 0.000) \); however 25% tested within the normal range of the HADS anxiety scale. Study authors suggest this was indicative of the time line in which the patient is asked to consider when completing the HADS (Whelan-Goodinson et al., 2009). Several studies have found that the HADS total score shows a higher correlation with depression and anxiety criterion measures than the subscale do (McDowell, 2006). Lisspers and colleagues (1997) found the correlation with the BDI was .71 for the HADS-depression subscale but for the total HADS the correlation with was .73. For hospital |
Evidence-Based Review of Moderate to Severe Acquired Brain Injury

Assessment of Outcomes Following Acquired/Traumatic Brain Injury

Outpatients the HADS-depression subscale correlated .77 with the Montgomery-Asberg Depression Rating (MADR) scale with a group of psychiatric patients (.70). Again with a group of elderly depressed patients the HADS and the MADR correlated .54 and .79. Overall, Mykletun and colleagues (2001) have reported the correlation between the sub-scores and the over all score as reliable.

**Discriminant Validity:** Correlation between the subscales of the HADS and the correlation between the HADS total score and other scales (the General Health Questionnaire (GHQ-28)) and the MADR scale can vary considerably. Aylard et al (1987) found the correlation of the two subscales of the HADS was $r = -.04$ compared to the subscale on the GHQ-28 was $r = .54$. Lewis and Wessely (1990) found the correlation between the HAD total score and the GHQ-28 was .75.

**Predictive validity:** The HAD-depression and anxiety subscales and were found to account for 52.6% and 60% (respectively) of variance when looking at patients who were diagnosed with a mood disorders and those with no psychiatric disorder (Herrero et al., 2003).

**Responsiveness**

In studies of primary care population, the HADS was successful in detecting DSM-III defined psychiatric morbidity, with the ROC curve showing a score of 8+ being optimal (Bjelland et al., 2006). When using the DSM III clinical interview schedule as the gold standard, ROC curves indicated $>/=9$ on the HADS anxiety subscale (sensitivity .66 and specificity .93) were indicative of caseness and scores of $>/=7$ on the HADS depression subscale (sensitivity .66 and specificity .97) were indicative of caseness. (Bjelland et al., 2006). Beck et al., (1997) found the HADS depression subscale had a AUR of .74 (SE =.09) with a cut off score of $>/=5$ yielding the highest efficiency at 72% with a sensitivity of 85%, but the specificity of only 47%. According to Herrero et al (2003), the curve ROC shows the model discriminates between cases and non cases: HAD-D (area-0.887; 95%CI: 0.84 to 0.91), HAD-A (area-.917; 95%CI: 0.88.to 0.95). For each of these two subcales the predicative power is 80% (HAD-D) and 83% (HAD-A). For the full scale the predicative power is 81%. Herrmann (1997) found the HADS correlated well with other quality of life indicators used in a variety of studies looking at patients with HIV, renal insufficiency, etc. the HADS anxiety subscale correlated well with chest pain, tachycardia, dizziness, etc. The HADS depression subscale correlated well with dyspnea, low exercise tolerance etc.

**Tested for ABI/TBI patients?** Yes the scale has been tested with an ABI population.

**Other Formats**
The scale has been translated into Arabic (Malasi et al., 1991), Dutch, French, German, Hebrew, Swedish, Italian and Spanish. All are available at no cost (Zigmond and Snaith, 1983). Recently a computer administered version using a touch screen has been developed and has been found to be as valid as the paper and pencil version (McDowell 2006).

**Use by Proxy?** The scale is designed to be completed by the individual.

**Advantages:** The HADS is brief and simple to use and although it was originally designed to be used with hospital populations it has been found to perform well with non-hospital groups (McDowell, 2006). It has been found to take on average 2-5 minutes to
Evidence-Based Review of Moderate to Severe Acquired Brain Injury | 2013

complete and is completed by the patients themselves (Snaith, 2003). The HADS requires the individual to respond to the question in relation to how they felt in the past week, thus it is therefore reasonable to re-administer the test again but only at weekly intervals. It has been found to perform as well as the BDI and the GHQ instruments. Overall, Mykletun et al. (2001) found the HADS scale possessed good “psychometric properties in terms of factor structure, intercorrelation, homogeneity and internal consistency” (p 543).

Limitations: When using the HADS to diagnosis depression or depressive symptoms post ABI, the sequelae of TBI may confound the test scores (Whelan-Goodinson et al., 2009). Caution is recommended when interpreting the results of these scales (Dawkins et al., 2006). Even though the HADS has been shown to be a reliable measure of emotional distress post ABI, the cut off scores and categories have not been shown to be useful in predicting probable presence or “caseness” of depression or anxiety (Whelan-Goodinson et al., 2009).

Summary – HADS
Interpretability: The results are easy to interpret with higher scores on each individual scale or the entire scale indicating greater anxiety, depression or mood disorders.
Acceptability: The HADS is widely accepted and used with most patient populations including those with a TBI.
Feasibility: The scale is readily available and can be used free of charge. It takes only a few minutes to complete, no specialized training is need to administer the test and may be completed by the patients themselves.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
<td>+++ (IC)</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>+++ (TR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver;

17.14 Mayo-Portland Adaptability Inventory (MPAI-4)
The Mayo-Portland Adaptability Inventory is based on an earlier scale: the Portland Adaptability Inventory (Lezak, 1987). Specifically designed for the evaluation of individuals during the post acute period following acquired brain injury, the scale was developed to provide a representation of the sequelae of ABI through the use key indicators of abilities, activities and social participation (Malec 2004a). Assessment with the MPAI is intended to yield information applicable to the development and ongoing evaluation of rehabilitation services within this population (Malec and Lezak 2003).
The original version of the MPAI consisted of six subscales: physical/medical, cognition, emotion, everyday activities, social behaviours and behaviours (Bohac et al. 1997). Items were rated to reflect distinctions between impairment, disability and handicap as defined by the World Health Organization’s ICIDH (Malec et al. 2000, Malec and Lezak 2003). The MPAI has undergone successive revisions based on ongoing Rasch and multivariate analyses. The most current version is the MPAI-4, which evaluates the general dimension of sequelae of ABI in 3 sub-dimensions: ability, adjustment and participation (Malec 2004).

The MPAI-4 consists of 29 items in 3 subscales (the Ability Index, the Adjustment Index and the Participation Index) plus an additional 6 items that are not included in the MPAI-4 score. The first 29 scale items are intended to reflect the current status of the individual with brain injury without attempting to determine whether their status might be influenced factors other than ABI. The additional six, unscored items are intended to identify the presence of other factors that may be contributing to the individual’s current status (Malec and Lezak, 2003).

In general, items are rated on a 5-point scale from 0–4, where 0 represents the most favourable outcome, no problem or independence and 4 represents the presence of severe problems. A worksheet is provided that guides the user through the scoring and re-scoring of items. Following any necessary re-scoring, item scores are summed for each subscale to provide a raw score for that index. After making adjustment for items appearing in more than one index, subscale raw scores are summed to provide an overall adaptability index score. Raw scores for the indices and total scale may be converted to T-scores with a mean of 50 and a standard deviation of 10 using the tables provided in the manual (Malec and Lezak, 2003). T-scores provided are based on data sets from two populations of individuals with ABI. They have not been referenced to non-ABI samples. In general, when compared to the reference populations with ABI, total T-scores less than 30 are indicative of good outcome, 30 – 40 of mild limitations, 40 – 50 of mild to moderate limitations, 50 – 60 of moderate to severe difficulties and above 60 of severe limitations (Malec and Lezak, 2003).

The MPAI-4 was designed to be completed by professional staff, individuals who have experienced brain injury and their significant others. Ratings provided by any 2 or more of these groups can be combined to provide a more comprehensive composite score (Malec and Lezak 2003). When administered by professional staff, the ratings should be completed by team consensus. The MPAI-4 is free of charge. The manual and rating forms may be downloaded from the COMBI website (http://tbims.org/combi/mpai). A French translation of the rating form is also available from the website.

<table>
<thead>
<tr>
<th>Table 17.34 Characteristics of the Mayo-Portland Adaptability Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
</tr>
</tbody>
</table>

www.abiebr.com

Updated August 2013
**Validity**

**Construct validity:** Principal components analysis of MPAI after Varimax rotation revealed 8 orthogonal factors, each with few items — factors corresponded to variables labeled Activities of Daily Living, Social Initiation, Cognition, Impaired Self-Awareness/Distress, Social skills/Support, Independence, Visuoperceptual, Psychiatric — several items loaded significantly on more than one factor (Bohac et al. 1997); principal component analysis of 22-item MPAI revealed five factors — one 8-item set was identified with acceptable levels of person separation and reliability — correlation between MPAI-30 item and MPAI – 22 item = 0.98 (Malec et al. 2000); item cluster analysis provided a 3-cluster solution that was substantially similar to the item groupings derived rationally by the scale authors — the cluster analysis solution was not statistically superior to the rational item groupings — factor analysis revealed 7 factors with eigenvalues > 1 though each factor contained few items — moderate correlations between subscales (0.49 to 0.65) suggested that subscales/dimensions may be assessing different aspects of a single underlying construct (Malec et al. 2003)

**Construct validity (known groups):** significant differences (p<0.001) in MPAI scores were identified in groups differentiated by Rancho Levels of Cognitive Functioning Scale (Malec & Thompson, 1994); staff-completed MPAI-22 ratings distinguished between patients receiving specialized vocational services (SVS), those receiving community reintegration services + SVS and those receiving comprehensive day treatment + SVS (p=0.0001; Malec et al. 2002)

**Concurrent validity:** original MPAI consensus ratings correlated with DRS scores (r=0.81), with Rivermead Behavioral Memory Test (r=-0.47) and with various neuropsychometric/cognitive measures (correlations ranged from 0.04 with WJRead to 0.56 with Stroop Color-Word test; Malec & Thompson, 1994)

**Predictive validity:** Pre-treatment MPAI-30 score was reported to be predictive of outcome post-treatment as assessed by the MPAI 30 (0.52), MPAI-22 (0.51), Goal Attainment Scaling (-0.49), the Independent Living Status (-0.32) and the Vocational Independence Scale (-0.26) — the pre-treatment MPAI-22 was similarly predictive of outcome, although the association with VIS scores was weak (-0.17) — at one year follow-up, the pre-treatment MPAI-30 and MPAI-22 scores were predictive of ILS and VIS scores (-0.25 & -0.34 vs. -0.26 & -0.32, for the 30 item and 22 item version, respectively) (Malec et al. 2000); pre-admission MPAI-22 score predicted independent living scale scores (concordance = 70.2%, p<0.01) and vocational independence scale scores (concordance = 66.9%, p<0.05) at 1 year follow-up following comprehensive day treatment (Malec, 2001); time since injury
and staff rated MPAI-22 were identified as significant predictors of Vocational Independence scale scores at job placement in the medical/vocational case coordination system (p<0.01), staff-rated MPAI-22 was also predictive of time to placement (p<0.001) (Malec et al. 2000b); staff MPAI-22 ratings contributed significantly to the prediction of community-based employment at one year follow-up (p<0.01; Malec et al. 2002)

### Responsiveness
MPAI provides a broader assessment at lower levels of disability than DRS (Malec & Thompson, 1994)
Change in MPAI-22 score from pre-admission to end of comprehensive day treatment program was significant (paired t= 8.35, p<0.0001; Malec 2001)

### Tested for ABI/TBI patients?
Specific to persons with acquired brain injury.

### Other Formats
**Mayo-Portland Participation Index (M2PI)**— the Participation Index from the MPAI-4, which may be used as a brief measure focused on participation based on indicators of community integration. Lower scores are indicative of greater community integration (Malec, 2004). Person reliability and separation were reported to be 0.78 & 1.89, respectively (Malec et al. 2003). In the same study, item separation = 7.59, item reliability = 0.98, and \( \alpha = 0.83 \). Reported person reliability & separation = 0.85 & 2.41 respectively and item reliability & separation = 0.99 & 8.17 respectively for staff-rated M2PI – person reliability for significant other and patient ratings as well as various composite ratings ranged from 0.74 (individual with ABI) to 0.89 (staff+ SO+ person with ABI) while item reliability ranged from 0.97 – 0.99 (Malec, 2004). When comparing ratings obtained from persons with ABI, significant others and staff, it was reported that persons with ABI tended to rate themselves as having greater independence and involvement in the community than raters from either of the other 2 groups. Overall agreement was greater for more concrete, functional items than for social indicators. No substantial floor or ceiling effects were reported – very high scores and very low scores were not common (<5% and <7% respectively) (Malec, 2004)

### Use by proxy?
Total MPAI scores derived by staff and significant others were correlated (r=0.47, p<0.005) as were scores derived by significant others and the patients themselves (r=0.37, p<0.025), but MPAI scores derived by staff and patients were not significantly correlated (r=0.09). Differences between staff and patient ratings are attributed to impaired self awareness (Malec et al. 1997). The authors speculated that differences between ratings may, in part, have been due to differing interpretations of terminology used in the test and differences in personal value assigned to various items.

Similar levels of reliability were obtained for ratings completed by staff, significant others and persons with ABI (Malec, 2004a). On the full MPAI-4, 42% of ratings made by staff, persons with ABI and their significant others (SO) were in complete agreement, however, a reliable difference in ratings was reported (rater reliability = 0.95). SO and staff raters tended to rate the person with ABI as more impaired than the person with ABI themselves did – this trend was observed for the adjustment and participation subscales, however, on the ability subscale individuals with ABI rated themselves as being more impaired than the staff members did. Exact agreement on subscale ratings were 41% for the Ability Index, 38% for the Adjustment Index and 46% for the Participation Index. More concrete items demonstrated greater agreement between rater groups, using techniques available via Facets analysis, ratings obtained from the profession care team, the individual with ABI and a significant other may be combined to provide a single composite score thereby providing a partial remedy to biases.
Advantages The MPAI is a readily available assessment of the post-acute sequelae of ABI. The Participation Index may be administered independently to provide a quick evaluation of participation outcomes. Differences in ratings between staff member consensus and individual with ABI or between SO and individual with ABI may provide a measure of impaired self-awareness (Malec et al. 1997; Malec & Degiorgio, 2002).

Limitations The authors do not recommended for the MPAI-4 for use in the assessment of individuals with very severe ABI (Malec & Lezak, 2003).

The authors reported that the placement of items in the 3 scale indices is based on a rational process in keeping with clinical observation and the results of ongoing analyses (Malec et al. 2003). However, the placement of some items appears odd. Self-care, for instance, is part of the participation index. In an earlier analysis, it was stated that it was more conceptually sound to place the self-care items with other basic skills such as use of hands, mobility and speech (Bohac et al. 1997). These basic items are currently part of the MPAI-4 abilities index. Other items, such as initiation, social contact and leisure skills/recreation were assigned to more than one index suggesting significant overlap between the subscales of adjustment and participation.

There are no published validation or reliability studies of the Mayo-Portland Adaptability that did not originate from the group responsible for the development of the scale.

Summary – MPAI
Interpretability: Tables are provided -- raw scores converted to standardized T-scores based on a national sample (n=386) or regional sample (n=134) – no truly normative data is available for the purpose of comparison.
Acceptability: May be completed by patients and significant others with trained professionals available to provide assistance.
Feasibility: The MPAI-4 is free to download and copy. Administration, scoring and interpretation should be undertaken by trained professionals. The manual also contains a recommendation that a person capable in advanced psychometrics should be available. To maintain high levels of reliability, assessment should be completed by team consensus.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+</td>
<td>+++ (IC)</td>
<td>+</td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; + = Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.
17.15 Medical Outcomes Study Short Form 36 (SF-36)

The Medical Outcomes Study Short Form 36 (SF-36) is a generic health survey created to assess health status in the general population as part of the Medical Outcomes Study (Ware & Sherbourne 1992). It is comprised of 36 items drawn from the original 245 items generated by that study (Ware & Sherbourne 1992; McHorney et al. 1993).

Items are organized into 8 dimensions or subscales; physical functioning, role limitations- physical, bodily pain, social functioning, general mental health, role limitations – emotional, vitality, and general health perceptions. It also includes 2 questions intended to estimate change in health status over the past year. These 2 questions remain separate from the 8 subscales and are not scored. With the exception of the general change in health status questions, subjects are asked to respond with reference to the past 4 weeks. An acute version of the SF-36 refers to problems in the past week only (McDowell & Newell 1996).

The recommended scoring system uses a weighted Likert system for each item. Items within subscales are summed to provide a summed score for each subscale or dimension. Each of the 8 summed scores is linearly transformed onto a scale from 0 – 100 to provide a score for each scale. In addition, a physical component (PCS) and mental component score (MCS) can be derived from the scale items. Standardized population data for several countries are available for the SF-36 (McDowell & Newell 1996). The component scores have also been standardized with a mean of 50 and standard deviation of 10 (Finch et al. 2002).

The SF-36 questionnaire can be self-completed or administered in person or over the telephone by a trained interviewer. It is considered simple to administer and takes less than 10 minutes to complete (Andresen & Meyers 2000). Permission to use the instrument should be obtained from the Medical Outcomes Trust who oversee the standardized administration of the SF-36 and will provide updates on administration and scoring (McDowell & Newell 1996). Various computer applications are available to assist in scoring the SF-36 including free Excel templates that can be downloaded from the internet (Callahan et al. 2005).

Table 17.36 Characteristics of the Medical Outcomes Study Short Form 36

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest reliability:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Brazier et al. (1992)</td>
</tr>
<tr>
<td></td>
<td>calculated correlation coefficients ranging from 0.6 (social functioning) to 0.81 (physical functioning). Mean differences ranged from 0.15 (social functioning) to 0.71 (mental health) with 91 – 98% cases falling into the 95% CI (constructed as per Bland &amp; Altman); lower values reported in stroke population of 0.28 (mental health) to 0.80 (social functioning) – reported substantial variability in individual responses, particularly for role limitations – emotional (Dorman et al. 1998); Brazier et al. (1996) reported r=0.28 (social functioning) to 0.70 (vitality) over a retest period of 6 months; Andresen et al (1999) reported ICC ranging from .052 (social functioning) to 0.80 (mental health).</td>
</tr>
</tbody>
</table>
health), ICC for physical summary scores = 0.82 and ICC = 0.79 for mental summary scores; r=0.79 and 0.78 (p<0.001) for the MCS and PCS respectively with the test taken at 6 months post-injury and 2-3 weeks later (Dikmen et al. 2001)

**Internal Consistency:** Brazier et al. (1992) \(\alpha \geq 0.80\) for all subscales but social functioning (\(\alpha = 0.73\)). Reliability coefficients = 0.74 (social functioning) – 0.93 (physical functioning); Anderson et al. (1996) reported \(\alpha\) of 0.6 (vitality) to 0.9 (physical functioning, bodily pain and role limitations – emotional). Four scales fell below 0.80; Brazier et al. (1996) reported \(\alpha \geq 0.80\) for all subscales but social functioning (0.56) and general health (0.66) – inter-item correlations \(\geq 0.73\) with the exception of social functioning (0.56) and general health (0.66); Essink-Bot et al. (1997) reported \(\alpha = 0.76\) (general health) – 0.91 (physical functioning); Hobart et al. (2002) found \(\alpha\) of 0.68 (general health) & 0.70 (social functioning) to 0.90 (physical functioning) -- Correlations between 8 scales were lower than the reported alpha coefficients; Hobart et al. (2002) found item-own exceeded item-other correlations by >2.5 SE for 6 of 8 scales - social functioning scale & general health scale did not (i.e. limited ability to distinguish constructs); Walters et al. (2001) reported \(\alpha \geq 0.80\) for all subscales but social functioning (\(\alpha = 0.79\); Doninger et al. (2003) reported person separation estimates of 2.27 and 2.35 for physical health and emotional health respectively; calibration of the physical functioning items generated a reliability of 0.84 with no misfits; calibration of the mental health and vitality scales yielded a reliability of 0.85 with one misfit; for all subscales \(\alpha\) ranged from 0.68-0.87 for controls, 0.83-0.91 for mild TBI and from 0.79-0.92 for moderate/severe TBI (Findler et al. 2001)

**Validity**

**Construct validity:** Walters et al. (2001) reported significant relationships in expected directions to support construct validity among older adults -- scores in all scales were reported to decrease as age increased (p<0.001) – women reported worse health than men on all scales even after adjusting for age (p<0.001) — respondents who had recently visited their physician reported poorer health on all scales (p<0.001) and people living alone had lower scores (p<0.001) except on general health (p=0.02); Doninger et al. (2003) reported item separation estimates of 12.03 and 7.95 for physical health and emotional health respectively; in a trauma population, principal components analysis revealed physical function, role physical and bodily pain had the strongest loadings on physical health and the lowest loadings on mental health whereas role emotional and mental health did the opposite – the general health, vitality, and social function scales had substantial loadings on both components – these results were comparable correlations found for the general US population (MacKenzie et al. 2002).

SF-36 scales correlated significantly with the Symptom Checklist (SCL), the Beck Depression Inventory (BDI-II) and the Health Problems List (HPL) – in the mild TBI group, scales related to physical functioning were strongly correlated with the HPL (-0.6 to -0.75) and the physical symptoms scale of the SCL (-0.5 to -0.63) – scales related to mental health were most strongly correlated with psychological factors on the SCL – strong correlations were found between BDI-II scores and all of the SF-36 scales, the highest with the mental health scale (-0.77) – in the moderate/severe group, correlations were weaker and more consistent but the strongest correlations were found where expected (Findler et al. 2001).

**Construct validity (known groups):** Patients diagnosed with \(\geq 1\) chronic physical problems, had lower scores on all dimensions of the SF-36 except mental health, than healthy age-matched controls(p<0.001). SF-36 scores distributed as expected for sex, age, social class and use of health services (Brazier et al. 1992); SF-36
Evidence-Based Review of Moderate to Severe Acquired Brain Injury

57

Module 17 - Assessment of Outcomes Following Acquired/Traumatic Brain Injury - V9-2013

www.abiebr.com

Updated August 2013

distinguished between groups based on functional dependence vs. independence based on BI scores (p<0.05 on all scales) and between groups based on mental health vs. ill-health defined by GHQ-28 scores (p<0.05 on all scales) (Anderson et al, 1996; Mayo et al. (2002) reported SF-36 scores discriminated stroke survivors from age and gender-matched controls; Williams et al. (1999) found the SF36 unable to discriminate between groups based on patient self-report ratings of overall HRQOL (same, a little worse or a lot worse than prestroke). SF-36 discriminated between age groups (< 75 yrs vs 75+) on physical functioning, vitality and change in health subscales (p ≤ 0.006) & between groups based on setting (general practice vs. hospital outpatients) on the physical function and role functioning – physical subscales (p=0.16) (Hayes et al. 1995); Essink-Bot et al. (1997) reported SF-36 able to discriminate between migraine sufferers and controls on all subscales (p<0.01; ROC/AUC = 0.54 – 0.67) and between groups of migraine sufferers based on absence from work (0 vs. ≥ 0.5 days; p<0.01, ROC/AUC = 0.61 – 0.79); Brazier et al. (1996) reported SF-36 scores distinguished groups based on recent visits to GP, hospital inpatient stays & longstanding illness (p<0.05); at 3 months and 1 year post-injury, mild TBI patients scored significantly lower than the matched normative group on all subscales and there was a significant negative correlation between number of post-concussion symptoms and SF-36 scores (Emanuelson et al. 2003); there were significant differences in scores between the control/nondisabled group, mild TBI group, and moderate/severe TBI – both TBI groups scored significantly lower than the control group on all scales and the mild TBI group scored significantly lower than the moderate/severe group on all scales except for the physical function subscore, which did not differ between TBI severity levels – after controlling for depression, many of the differences between the 2 TBI groups became insignificant (Findler et al. 2001); the self-ratings of matched-normal controls were found to be significantly higher than those of TBI patients on all scales except for the general health scale – the PCS and MCS also differed significantly between controls and TBI patients (Paniak et al. 1999)

Construct validity (convergent/divergent): Correlations of −0.41(social functioning vs. social isolation) to −0.68 (vitality vs. energy) between similar scales on the SF-36 and the Nottingham Health Profile were reported. Correlations between dimensions less clearly related ranged from −0.18 (physical functioning vs. emotional reaction) to −0.53 (social functioning vs. emotional reactions) (Brazier et al. 1992); Anderson et al. (1996) reported BI scores (in stroke survivors) strongly associated (p<0.001) with physical functioning & general health -Mental health on the GHQ28 most strongly associated (p<0.001) with the social functioning, role limitations – emotional and mental health scales of the SF-36; Dorman et al. (1999) reported SF-36 physical functioning subscale correlated most closely with mobility, self-care and activities domains of EuroQol (r=0.57, 0.65 & 0.63) and less strongly with the EuroQol psychological domain (0.34) – SF-36 bodily pain correlated with EuroQol pain domain (r=0.66) and moderately with all EuroQol domains – role functioning, emotional correlated most closely with EuroQol psychological domain (r= 0.43) and least with EuroQol self care (r=0.24)- SF-36 mental health was not closely related to the psychological domain (r=0.21) or to physical EuroQol domains (r=0.06 – 0.10). – SF-36 general health correlated with EuroQol overall HRQOL rating) r=0.66; Lai et al. (2003) reported r=0.55 between SF-36 physical functioning scale and BI; Andresen et al. (1999) reported physical health scores correlated more strongly with ADL scores than with GDS (-0.38 vs. − 0.28) and mental health summary scores correlated more strongly with GDS scores than ADL scores (-0.63 vs. 0.01) – however, role- physical correlated more strongly
with GDS scores than ADL scores, contrary to a prior hypothesis – social functioning, role-emotional, vitality and mental health all correlated more strongly with GDS scores than ADL scores; Dikmen et al. (2001) found significant correlations between the PCS and the Functional Status Examination (FSE) regardless of whether the patient (-0.68) or a significant other (-0.64) assessed patient function – the correlations between the MCS and the FSE were weak and nonsignificant; McNaughton et al. (2005) reported high correlations (0.32-0.97) across the PCS, FIM, Barthel Index and the London Handicap Score – correlations of these measures with the MCS were less strong (0.17-0.32)

**Predictive validity:** McHorney (1996) examined data from Medical Outcomes Study – reported general health perceptions scale to be most predictive of death (death rate of patients in lowest quartile for SF-36 general health scale was 3 times greater than for patients with SF-36 scores in the highest quartile), followed by scores in physical functioning. Baseline physical functioning, role functioning-physical and pain scales were most predictive of hospitalizations and pain, general health and vitality were most predictive of physician visits.

### Responsiveness

Via item mapping -- social functioning subscale limited assessment of number & difficulty of activities – demonstrated marked ceiling effects – up to 60% for MRS grade 0) SF36 physical function scale reported to have floor effects of 37% & 100% for patients with MRS grades 4 & 5 (Lai et al. 2003); Large ceiling effects reported for the role limitations – physical (53%), bodily pain (43%), social functioning (67%) and role limitations – emotional scales (72%) – no floor effects over 7% were reported -- scores for SF-36 physical functioning scale more uniformly distributed than BI scores suggesting lower floor & ceiling effects than the BI. (Anderson et al. 1996; Brazier et al. (1996) reported floor effects in excess of 25% for role limitations physical and emotional and ceiling effects >25% for social functioning and role limitations emotional & physical.

Notable floor effects (role limitations – physical 59.1%; role limitations – emotional 19.9%) and ceiling effects (role limitations – emotional 63.1%; social functioning 29.9%; bodily pain 25.6%) reported among ischemic stroke survivors (Hobart et al. 2002); Substantial floor and ceiling effects reported by O’Mahoney et al. (1998); For face-to-face, telephone and self-administration, Weinberger et al. (1996) reported substantial floor effects for role-physical (> 40%) and role –emotional (>25%) subscales & ceiling effects for role-emotional (>36%) and social functioning subscales (>27% - for face-to-face and self administration only); Walters et al. (2001) reported substantial floor (30.9 – 61%) & ceiling effects across all age groupings (65-69, 70–74, 75-79, 80 – 84 & 85+) in the role functioning physical (30.9% – 61% & 11.7% – 38.6%) and role functioning – emotional (25.6% - 50.4% & 32.2% - 53.2%) as well as substantial ceiling effects in social functioning and bodily pain (15% – 46.7% & 14.1% - 21.1%, respectively); Andresen et al. (1999) reported substantial floor effects of 26.8% and 29.5% for physical functioning and role-physical, respectively, in a sample of nursing home residents as well as ceiling effects of 36.1%, 49.5% and 21.6% in social functioning, role-emotional, and bodily pain respectively.

Mossberg & McFarland (2001) found SF36 effect sizes (admission to outpatient rehabilitation to discharge) = 0.48 (role limitations – emotional) to 1.38 (bodily pain) - PCS (physical component) & MCS (mental component) effect sizes = 0.80 and 0.45 respectively.

Effect sizes for the PCS and MCS were 2.48 and 0.93 respectively (Paniak et al.)
Yes, several studies have been published indicating the scale has in fact been tested with those who have sustained a TBI. (Ocampo & Dawson 1997, Doninger et al. 2003, Emanuelson et al. 2003, Findler et al. 2001; Dikmen et al. 2001; Brown et al. 2004; Paniak et al. 1999; Corrigan et al. 1998; MacKenzie et al. 2002; Callahan et al. 2005; McNaughton et al. 2005)

Other Formats

**Mailed questionnaire**: Hayes et al. (1995) found type/mode of administration was clearly related to completeness of data (p<0.0001). For self-completion vs. in-person interview, percentage of missing items greater among the older respondents (p<0.015). Time to complete survey not dependent upon mode of administration or age -- 84% of the respondents completed the assessment in 10 minutes or less; Walters et al. (2001) reported non-completion of the mailed survey to be significantly related to increasing age (p<0.001).

**Face-to-face, self report and telephone interview**: Weinberger et al. (1996) reported internal consistency for all modes of administration – face-to-face $\alpha = 0.75 - 0.89$; self $\alpha = 0.77 - 0.93$; telephone $\alpha = 0.67 - 0.92$. Mean test-retest correlations for face-to-face, self & telephone modes were 0.80, 0.83 and 0.79. Between mode correlations were similar – face-to-face vs. self $r=0.54 - 0.82$; face-to-face vs telephone $r=0.55 - 0.91$. Correlations did not differ significantly by order of administration. Despite short testing intervals, large absolute differences were reported on within mode and between mode comparisons. Directional differences (over time < 1 week) were significant on between mode comparisons on 4/8 subscales (physical function, social function, role-emotional & mental health) with face-to-face interviews producing higher scores.

**Acute (1-week recall) version**: Keller et al. (1997) reported median inter-item correlations ranged from 0.43 (role-emotional) to 0.78 (bodily pain) -- $\alpha$ ranged from 0.59 (role-emotional) to 0.89 (physical functioning). Vitality, role emotional and mental health $\alpha$ values fell below 0.80. Principal component analysis revealed the same 2 factor structure as the standard version. The acute version displayed significant ceiling effects (>20%) in 4 subscales (role-physical, bodily pain, social functioning & role-emotional). There were no reported floor effects. Change scores for the acute form (baseline to week 4) were more closely related to one-week change in disease severity than standard form scores. For acute change scores 10/18 such comparisons reached significance.

**Proxy Assessment**

Dorman et al. (1998) reported test-retest reliability better when the patient completed the forms than when completed by proxy respondent. ICC’s ranged from 0.3 (mental health) to 0.81 (bodily pain/general health) when forms were patient-completed vs ICC of 0.24 (mental health) to 0.76 (social functioning) for proxy completion.

Pierre et al. (1998) demonstrated poor to moderate agreement between proxy and patient ratings. In a rehabilitation setting, ICC’s = 0.01 (social functioning) to 0.60 (vitality) for patient/health professional proxy pairings. For significant others proxies/patients, ICC’s = -0.11(mental health) – 0.58 (general health). In a day hospital setting & professionals as proxies, ICC’s = 0.09 (role physical) – 0.45 (physical functioning) -- with sig. others, ICC’s = 0.01 (social functioning) – 0.71 (physical functioning). $\alpha = 0.64 – 0.86$ for the patient data, 0.76 – 0.90 for the health professional data and 0.69 – 0.84 for the significant other data.

Segal & Schall (1994) reported ICC of 0.15 (role limitations – emotional) to 0.67 (physical functioning) for patient ratings vs. proxy ratings.
Ocampo and Dawson (1997) found the highest level of agreement between TBI patients and their informants was for physical functioning (ICC=0.58) and general health (ICC=0.51); agreement for role-physical and role-emotional were high for the moderate and severe groups; agreement was generally poor on the other subscales.

Dikmen et al. (2001) reported a correlation of 0.53 (p<0.001) between the assessments of patients and their significant other on the PCS, but this correlation on the MCS was weak and non-significant.

**Advantages.** The SF-36 is simple to administer. Either of the forms, the self-completed or the interview form, takes less than 10 minutes to complete (Hayes 1995). As a self-completed, mailed questionnaire, it has been shown to have reasonably high response rates: 83% has been reported by Brazier et al. 1992 and O’Mahoney et al. 1998; 75% - 83% reported by Dorman et al. 1998; in 1999 Dorman and colleagues reported a response rate of 85% and in 2001 and Walters et al. (2001) reported 82% overall & 69% for those over age 85.

Callahan et al. (2005) found that the SF-36 was appropriate for longitudinal serial assessment of recovery in a mixed group of patients suffering from one of cerebrovascular accident, traumatic brain injury, or spinal cord dysfunction. The instrument has been shown to be valid and reliable in the adult TBI population and appears to be sensitive to the wide spectrum of health issues faced by this group (Emanuelson et al. 2003; Findler et al. 2001).

**Limitations.** Higher rates of missing data have been reported among older patients when using a self-completed form of administration (Brazier et al. 1992; Hayes et al. 1995; Brazier et al. 1996). O’Mahoney et al. (1998) found item completion rates to range from 66% to 96%. At the scale level, complete data collection (amount required to compute a scale score) ranged from 67% (role limitations – emotional) to 97% (social functioning). Walters et al. (2001) reported scale completion rates among community dwelling older adults ranging from 86.4% to 97.7% with all eight scales being calculable for 72% of respondents. Dorman et al. (1999) reported proportion of missing data on the scale level ranging from 2% (social functioning) to 16% (role functioning – emotional). Given the lack of data completeness found, postal administration of the SF-36 may not be appropriate for use among older adults. However, low completion rates may not be limited to self-completion or postal administration. Andresen et al (1999) administered the SF-36 to nursing home residents by face-to-face interview and reported that only 1 in 5 residents were able to complete it.

It has been suggested that data completeness may be indicative of respondent acceptance and understanding of the survey as relevant to them (O’Mahoney et al. 1998; Andresen et al. 1999). Hayes et al. (1995) noted that the most common items missing on the self-completed questionnaire referred to work or to vigorous activity.
Older respondents identified these questions as pertinent for much younger people and not relevant to their own situation. The authors suggested modifications to some of the questions, which may increase acceptability to older populations. In a qualitative assessment of the physical functioning and general health perceptions dimensions of the SF-36, Mallinson (2002) noted that the participants, who were all over the age of 65, tended to display signs of disengagement from the interview process and some participants expressed concern relating to the relevance of the questions. There was also considerable variation noted in subjective interpretation of items and most subjects used qualifying, contextual information to clarify their responses to the interviewer. As Mallinson pointed out, individual issues of subjective meaning and context are lost when the questionnaire is scored.

The SF-36 does not lend itself to the generation of an overall summary score. In scales using summed Likert scales, information contained within individual responses is lost in the total scale score (ie. any given total score can be achieved in a variety of ways from individual item responses) (Dorman et al. 1999). Hobart et al. (2002) examined the use of the 2-dimensional model, which consists of a mental health component (MCS) and physical health component (PCS). These two scales can account for only 60% of the variance in SF-36 scores suggesting a significant loss of information when the 2-component model is used.

It has been suggested that the SF-36 may be more sensitive to the health difficulties of mild TBI than of moderate/severe TBI patients as it was unable to differentiate between the severity levels (Emanuelson et al. 2003). One study found initial differences between these groups, however once depression was controlled for these differences were less visible, suggesting that depression may account for the differences between TBI groups on the SF-36 (Findler et al. 2001). MacKenzie et al. (2002) suggest that adding a cognitive component to the SF-36 would make the instrument a more useful outcome measure in a head trauma population, as the tool is likely to underestimate the extent of disability in this group.

The level of test re-test reliability reported in stroke populations indicate that the SF-36 may not be adequate for serial comparisons of individual patients, but rather should be used for large group comparisons only (Dorman et al. 1998). Weinberger et al. (1996) also questioned the usefulness of the SF-36 in serial evaluation of individuals given large reported absolute differences in SF-36 scores obtained via common modes of administration (face-to-face interview, self administration and telephone interview) over short testing intervals.

Dikmen et al. (2001) emphasized that the SF-36 was designed to be self-administered, thus its disadvantage is the inability to use the SF-36 to assess patients who are too impaired to complete the questionnaire on their own. While the use of a proxy may be the only means by which to include data from more severely affected TBI patients,
reported disagreement between patient and proxy assessments has been considerable. In an adolescent TBI population, moderate rates of agreement were reported between proxy and patient respondent ratings for items related to physical health; however, on more subjective items, agreement was very low (Ocampo & Dawson 1997). It has been suggested that clinicians do not substitute proxy data for patient responses due to the subjective nature of many SF-36 items (Ocampo & Dawson 1997).

**Summary – MOS-SF36**

**Interpretability:** Use of scale scores and summary component scores represents a loss of information and decreases potential clinical interpretability. Standardized norms for several countries are available for the SF-36.

**Acceptability:** Completion times are approximately 10 minutes for either self-completed or interview administered questionnaires. Some items have been questioned for their relevance to elderly populations. The SF-36 has been studied for use by proxy, however, agreement rates are low and reliability of the test decreased when proxy respondents completed assessments.

**Feasibility:** The SF-36 questionnaire can be administered by self-completion questionnaire or by interview (either on the telephone or in-person). It has been used as a mail survey with reasonably high completion rates reported, however, data obtained are more complete when interview administration is used. Permission to use the instrument and additional information regarding its administration and scoring should be obtained from the Medical Outcomes Trust.

**Table 17.37 SF-36 Evaluation Summary**

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
<td>++ (TR)</td>
<td>+++</td>
</tr>
<tr>
<td>++ (IC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver; varied (re. floor/ceiling effects; mixed results)

**17.16 Mini Mental Status Examination (MMSE)**

The Mini-Mental State Examination was developed as a brief screening tool to provide a quantitative assessment of cognitive impairment and to record cognitive changes over time (Folstein et al. 1975). While the tool’s original application was the detection of dementia within a psychiatric setting, its use has become widespread.

The MMSE consists of 11 simple questions or tasks. Typically, these are grouped into 7 cognitive domains; orientation to time, orientation to place, registration of three words, attention and calculation, recall of 3 words, language, and visual construction. Administration by a trained interviewer takes approximately 10 minutes. The test yields a total score of 30 and provides a picture of the subjects present cognitive performance based on direct observation of completion of test items/tasks. A score of 23/24 is the
generally accepted cutoff point indicating the presence of cognitive impairment (Dick et al. 1984). Levels of impairment have also been classified as none (24-30); mild (18-24) and severe (0-17) (Tombaugh & McIntyre 1992).

An expanded version of the MMSE, the modified mini-mental state examination (3MS) was developed by Teng & Chui (1987) increasing the content, number and difficulty of items included in the assessment. The score of the 3MS ranges from 0 – 100 with a standardized cut-off point of 79/80 for the presence of cognitive impairment. This expanded assessment takes approximately 5 minutes more to administer than the original MMSE.

<table>
<thead>
<tr>
<th>Table 17.38 Characteristics of the Mini Mental State Examination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td>In an extensive review, Tombaugh &amp; McIntyre (1992) reported moderate to high <strong>test-retest reliability</strong> citing correlations of 0.38 to 0.99 in studies having a retest interval of &lt; 2 months (24/30 studies r &gt;0.75).</td>
</tr>
<tr>
<td><strong>Interobserver reliability:</strong> Molloy and Standish (1997) reported an ICC of 0.69 for the traditional MMSE; Dick et al. (1984) K = 0.63; concordance correlation coefficient = 0.87 between evaluations performed by GPs &amp; those performed by psychologists (Fabrigoule et al. 2003)</td>
</tr>
<tr>
<td><strong>Internal consistency:</strong> Cronbach’s α coefficient of 0.54 – 0.96 have been reported by Tombaugh &amp; McIntyre (1992)</td>
</tr>
<tr>
<td><strong>Validity</strong></td>
</tr>
<tr>
<td><strong>Concurrent Validity:</strong> Tombaugh &amp; McIntyre (1992) reported correlations of 0.70 to 0.90 between MMSE scores and other measures of cognitive impairment</td>
</tr>
<tr>
<td><strong>Construct validity</strong> - correlations between ADL scores and the MMSE of 0.40 – 0.75. (Tombaugh &amp; McIntyre, 1992) support importance of cognitive status to functional outcome. Grace et al. (1995) reported significant association between FIM scores and MMSE scores (p&lt;0.05); Agrell &amp; Dehlin (2000) reported significant correlations between MMSE scores and BI, MADRS and Zung Depression Scale (p&lt;0.05); demonstrated expected lower MMSE scores, stroke patients vs. controls (p&lt;0.001) – factor analysis revealed 3 factors explaining 53% variance; the MMSE showed strong correlations with the WAIS – verbal (r=0.78) and performance- IQ (r=0.66) scores (Folstein et al. 1975); Dick et al (1984) reported r=0.55 and r=0.56 for verbal &amp; performance IQ, respectively.</td>
</tr>
<tr>
<td><strong>Construct validity (known groups):</strong> MMSE scores could discriminate between groups based on categories of vocational recommendations (return to work, vocational training, supported work and continued remedial therapy; p&lt;0.0001), MMSE scores accounted for 36% variance between cell means (Mysiw et al. 1989); DePaolo and Folstein (1978) reported the MMSE able to distinguish between patients with cerebral abnormalities and those with peripheral disorders only (p&lt;0.0005)</td>
</tr>
<tr>
<td><strong>Predictive validity:</strong> Ozdemir et al. (2001) reported relationships between baseline MMSE scores &amp; change in motor-FIM from admission to discharge among stroke rehabilitation patients (r=0.31; p&lt;0.04), suggesting MMSE baseline scores are somewhat predictive of functional improvement.</td>
</tr>
<tr>
<td><strong>Sensitivity &amp; Specificity:</strong> Tombaugh &amp; McIntyre (1992) reported an average <strong>sensitivity</strong> of 75% among dementia patients. Among general neurology and psychiatry patients, sensitivity was lower; ranging from 21 – 76%. - major variable</td>
</tr>
</tbody>
</table>
in sensitivity was level of impairment; sensitivity of the MMSE increased with level of impairment. Low level of sensitivity is supported (Dick et al. 1984 – not sensitive to changes in patients with right-sided disease – not useful in discriminating focal vs. diffuse disease) particularly among stroke patients (Grace et al. 1995 – sensitivity 44%, area under curve = 0.7097; Agrell & Dehlin, 2000 – 56%). Agrell & Dehlin (2000) reported MMSE could discriminate between patients with left-sided and infratentorial lesions (p<0.05) though not between right-sided and left-sided lesion groups.

Tombaugh & McIntyre (1992) reported specificity of 62% - 100%; 80% (Agrell & Dehlin, 2000); 84% (Grace et al. 1995).

Blake et al., 2002 reported sensitivity = 62% and specificity = 88% in group of stroke patients no suitable cut-off point could be identified if MMSE is used as a screening measure for verbal or visual memory deficits

<table>
<thead>
<tr>
<th>Responsiveness</th>
<th>N/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested for ABI patients?</td>
<td>Mysiw et al. (1989) reported that the MMSE was able to distinguish between TBI patients classified by vocational recommendations. Keith et al. (1998) have used the MMSE as the tool against which the Cognitive Drug Research system was validated for use among brain injured patients, however, the MMSE itself has not, apparently, undergone a similar evaluation in this specific population.</td>
</tr>
</tbody>
</table>

| Other Formats | Modified Mini-mental State Examination (3MS): Grace et al. (1995) compared the MMSE directly to 3MS. The test-retest stability of the 3MS was reported as r=0.80; p<0.001. Concurrent/Construct Validity - the 3MS correlated strongly with the MMSE at admission and discharge (r=0.84 & 0.85, respectively; p<0.001); correlated with a battery of neuropsychological assessments (COWA, BNT, HVOT, LM, VR – immediate and delayed, WMS-R, LM immediate & delayed); association with functional outcome (FIM) -- stronger for the 3MS than for the MMSE (t=3.28; p<0.05). Using the standardized cut-off points for cognitive impairment & ROC analysis, the 3MS showed greater sensitivity than the MMSE (69% vs. 44%) and similar specificity (80% vs. 79%) – area under the curve – 0.7977 for 3MS. 3MS + Clock-drawing: To increase 3MS sensitivity among patients with right hemisphere stroke, Suhr & Grace (1999) advocate the addition of the Wilson clock-drawing test. A clock-drawing task added < 2 min. to administration; increased sensitivity among stroke patients with right hemisphere lesions (87%). This testing format maintained a strong association with FIM scores (p<0.005). Standardized MMSE (SMMSE): Molloy & Standish(1997) developed detailed instructions for administration and scoring of each item. Test retest variance was reduced by 86% and interobserver variance by 76% when the standardized MMSE was used. (SSMSE - ICC = 0.90; MMSE- ICC = 0.69). Telephone Version – ALFI-MMSE: Includes 22/30 of the original MMSE items, the majority of which were removed from the last section (language and motor skills). Correlations between phone and face-to-face versions = 0.85 (p<0.0001). Patients tended to do slightly better on in-person testing than on the telephone. Sensitivity (using a brief neurological screening test as the criterion) of 67% and specificity of 100% were reported in a population of elderly, community-dwelling individuals. This was similar to the sensitivity/specificity reported for screening with the traditional MMSE (68%; 100%). (Roccaforte et al. 1992) T-MMSE (26 item version of the ALFI-MMSE, Roccaforte et al. cited in Newkirk et al., 2004): T-MMSE correlated with the MMSE (r = 0.88; p<0.001), neither hearing impairment nor years of education were associated with T-MMSE scores. |
22 points in common between the 2 scales, scores were correlated ($r=0.88$ $p<0.001$), however, telephone scores tended to be higher than in-face scores ($p<0.01$) (Newkirk et al, 2004). The authors provide tables for the conversion of T-MMSE scores to MMSE scores.

**Use by Proxy?**

N/A

**Advantages.** The Mini-mental State Examination is brief, inexpensive and simple to administer. Its widespread use and accepted cutoff scores increase its interpretability.

**Limitations.** It has been suggested that the MMSE may attempt to assess too many functions in one brief test. An individual’s performance on individual items or within a single domain may be more useful than interpretation of a single score. (Wade 1992;Tombaugh & McIntyre 1992). However, an acceptable cut-off for the identification of the presence of an impairment may be possible only when the test is used as a measure of “cognitive impairment” (Blake et al. 2002). Blake et al. (2002) reported that when the test is used to screen for problems of visual or verbal memory, orientation or attention acceptable cut-off scores could not be identified.

MMSE scores have been shown to be affected by age, level of education and sociocultural background (Tombaugh & McIntyre, 1992; Bleeker et al. 1988; Lorentz et al. 2002). These variables may introduce bias leading to the misclassification of individuals. Though perhaps the prevalent view, such biases have not always been reported. For instance, Agrell & Dehlin (2000) found neither age nor education to influence scores. Lorentz et al. (2002) expressed concern that adjustments made for these biases may limit the general utility of the MMSE.

Perhaps the greatest limitation of the MMSE is its low reported levels of sensitivity particularly among individuals with mild cognitive impairment (Tombaugh & McIntyre, 1992; de Koning et al. 1998), in patients with focal lesions, particularly those in the right hemisphere (Tombaugh et al. 1992), within a general neurological patient population (Dick et al. 1984) and within a stroke population (Suhr & Grace 1999; Blake et al., 2002). It has been suggested that its low level of sensitivity derives from the emphasis placed on language items and a paucity of visual-spatial items (Tombaugh et al. 1992; Grace et al. 1995; de Koning et al. 1998; Suhr & Grace, 1999; de Koning et al. 2000). Various solutions have been proposed to the problem of the MMSE’s poor sensitivity including the use of age-specific norms (Bleecker et al. 1988) and the addition of a clock-drawing task to the test (Suhr & Grace, 1999). Clock-drawing tests themselves have been assessed as acceptable to patients, easily scored and less affected by education, age and other non-dementia variables than other very brief measures of cognitive impairment (Lorentz et al. 2002) and would have little effect on the simplicity and accessibility of the test. The MMSE has been thoroughly evaluated for use among a variety of neurological populations. Unfortunately, at present, information regarding the reliability and validity of the MMSE when used among patients with TBI/ABI is extremely limited.
Summary – MMSE

Interpretability: The MMSE is widely used and has generally accepted cut-offs scores indicative of the presence of cognitive impairment. Documented age and education effects have led to the development of stratified norms (Ruchinskas & Curyto 2003).

Acceptability: The test is brief requiring approximately 10 minutes to complete. It may be affected by such patient variables as age, level of education and sociocultural background. As it is administered via direct observation of task completion, it is not suitable for use with a proxy respondent.

Feasibility: The test requires no specialized equipment and little time, making it inexpensive and portable. A survey conducted by Lorentz et al. (2002) revealed participant physicians found the MMSE too lengthy and unable to contribute much useful information.

Table 17.39 MMSE Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
<td>+++ (TR)</td>
<td>+++</td>
</tr>
<tr>
<td>++ (IO)</td>
<td>++ (IC)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver; varied (re. floor/ceiling effects; mixed results)

17.17 Neurobehavioral Functioning Inventory (NFI)
The Neurobehavioral Functioning Inventory (NFI) was originally developed as part of the General Health and History Questionnaire, which was used to collect a variety of information on individuals who had experienced a traumatic brain injury (Kreutzer et al. 1987). The NFI is intended to assess a wide spectrum of behaviours and symptoms encountered in everyday life in order to evaluate the neurological, behavioural and psychological effects of head injury (Kreutzer et al. 1996, Weinfurt et al. 1999, Seel et al. 1997).

The NFI consists of 70 items representing behaviours or symptoms. These are grouped into 6 functional domains or subscales derived from principal components and factor analytic methodologies (Seel et al. 1997, Hart et al. 2003). The 6 domains are depression (13 items), somatic (11 items), memory/attention (19 items), communication (10 items), aggression (9 items) and motor (8 items)(Kreutzer et al. 1996, Hart et al. 2003). Six additional, critical items relating to patient safety and community integration have been added to the scale (Kreutzer et al. 1999) to be used in the identification of areas requiring immediate attention (Awad, 2002).

Items are rated for frequency of occurrence on a 5-point Likert scale from 1 (never) to 5 (always). While the NFI is a self-rating inventory, it provides for the inclusion of information obtained from suitable proxy sources. The test contains forms for ratings by
self and by a significant other. The test takes approximately 20 minutes to complete (Awad 2002)

The NFI is a proprietary scale that must be purchased from The Psychological Corporation (Harcourt Assessment, Inc.).

Table 17.40 Characteristics of the Neurobehavioral Functioning Inventory

| Reliability | Internal consistency: $\alpha$ values for each scale were reported to be 0.93 (depression), 0.86 (somatic), 0.95 (memory/attention), 0.88 (communication), 0.89 (agression), and 0.87 (motor impairment) – for the entire scale, $\alpha$ = 0.97 (Kreutzer et al. 1996); Awad (2002) reported $\alpha$ coefficients for each NFI subscale – depression = 0.93, somatic = 0.83, memory/attention = 0.95, communication = 0.88, aggression = 0.87 and motor = 0.88 |
| Validity | Construct validity: - factorial analysis of the original 105 scale items revealed a 70-item, 6 factor model with a comparative fit index of 0.89 that was superior to other models tested - intercorrelations between total subscale scores ranged from 0.44 to 0.67; Awad (2002) reported a Goodness-of-Fit index of 0.71 and Comparative Fit Index of 0.71 for the six factor, 70-item NFI – in general, fit indices for each subscale were higher than for the total scale -- 20 items had squared multiple correlations < 0.40 (1-depression, 9-somatic, 4-memory, 3-communication, 2-aggression and 1-motor item) – intercorrelations between subscales ranged from 0.56 – 0.58 and were all significant (p<0.001), suggesting that the NFI may be measuring a single, large underlying construct. Construct validity (known groups): - scores on depression (p<0.002), memory/attention (p<0.002), communication (p<0.001), aggression (p<0.002) and motor (p<0.002) subscales could distinguish between groups based on employed vs. unemployed persons who had sustained traumatic brain injury (Sander et al. 1997); comparison of subscale scores for individuals with TBI vs. non-clinical controls via ANCOVA revealed no significant differences between groups on the depression, memory/attention, communication and motors subscales when controlling for the effects of age and sex – the only significant differences appeared on the somatic subscale (p<0.01) on which non-clinical controls achieved higher scores than the TBI group (Awad, 2002) Concurrent validity: Scores on NFI subscales were correlated with the following scales from the MMPI – hypochondriasis, depression, hysteria, psychasthenia, schizophrenia. Correlations between NFI subscales and MMPI scales were all significant (p<0.001) – correlations between MMPI hypochondriasis and NFI subscales ranged from 0.24 (agression) and 0.65 (somatic), for MMPI depression correlations ranged from 0.21 (agression) and 0.47 (depression, motor and somatic), for MMPI hysteria from 0.25 (communication) to 0.50 (somatic), for MMPI psychasthenia from 0.26 (communication) to 0.43 (depression) and for MMP schizophrenia from 0.25(agression) to 0.40 (depression) (Kreutzer et al. 1996) NFI Communication subscale correlated significantly with scores on neuropsychological measures of attention, memory and learning, communication and visual and motor functioning (p<0.001) – no other subscale correlated significantly with any of the neuropsychological tests administered with the exception of Memory/Attention which correlated with scores on the Symbol Digits Modalities Test Oral (Kreutzer et al. 1996). NFI Digits Modalities Test Oral correlated significantly with WMS-Logical Memory raw
scores (r=-0.26, p<0.001) and with the WMS-R Logical Memory recall scores (r=-0.26, p<0.001), NFI motor scores correlated with Trailmaking tests A (r=0.25, p<0.001) and B (r=0.27, p<0.001) and Grooved Pegboard scores (r=-0.28, p<0.001), NFI communication correlated with Controlled Word Association Test adjusted scores (r=-0.18, p<0.001) (Awad, 2002)

**Responsiveness**
N/A

**Tested for ABI/TBI patients?**
Head injury specific.

**Other Formats**
- **NFI-66** – developed by Kreutzer & Devany (unpublished). Weinfurt et al. (1999) performed factor analysis revealing 4 components with eigenvalues > 2.0; cognitive deficits, depression, aggression and somatization. Internal reliability of the 4 scale NFI-66 ranged from 0.79 (aggression) to 0.92 (cognitive deficits and depression). Significant correlations were reported between NFI-66 scale scores and the GOS ranging from 0.21 (depression) – 0.26 (somatization). Aggression subscale scores did not correlate with GOS scores. Scores on the Euroqol VAS were significantly and inversely correlated with NFI subscale scores ranging from 0.17 (aggression) to 0.50 (depression)
- **NFI-D** (13-item Depression subscale of the NFI); Seel & Kreutzer (2003) reported high internal consistency (α=0.93). Convergent and discriminant validity was supported as scores on the NFI-D correlated with both Beck Depression Inventory scores (r=0.765) and MMPI-2 Depression scale T-scores (r=0.752) but not significantly with MMPI-2 hypomania scale scores (r=0.159). Normal and clinically depressed BDI scores were accurately predicted by NFI-D scores 81% & 87% of the time, respectively. Patients who were classified with mild or borderline depression on the BDI were less likely to be correctly classified as such by the NFI-D. Using the MMPI-2 Depression score classifications – normal vs. depressed classifications could be accurately predicted by NFI-D scores 75% & 83% of the time, respectively. Via mapping to the BDI, the following score ranges were proposed for the identification and classification of depression: ≤ 28 (minimal depression), ≥43 (clinical depression – moderate to severe), 29 – 42 (mood disturbance, however, classification in this borderline region contained many false positives and false negatives). Seel et al. (1997) reported agreement between family and patient ratings ranged from 48% to 84% and, for the most part, family members and patients tend to rate problems as occurring at the same frequency. On an item by item analysis, there were no statistical differences for ratings on 57 of 70 items.

**Use by proxy?**
Test contains forms for ratings by self and by significant other (proxy).

Correlations between self and SO ratings were moderate for communication and memory/attention and weaker/nonsignificant for motor, depression, somatic and aggression scales. Self ratings were significantly higher than SO ratings form somatic, memory/attention and communication scales (Rush et al. 2004)

Concordance coefficients (between patient and significant other) ranged from 0.63 (aggression) to 0.76 (somatic). Significant others rated symptoms in the aggression scale as being significantly more frequent than the patients. There was a trend identified for ratings of symptoms on the depression subscale to differ in the same manner. Such discrepancies were noted more for cognitive or behavioural symptoms not for physical or somatic ones (Hart et al. 2003).

Seel et al. (1997) reported agreement between family and patient ratings ranged from 48% to 84% and, for the most part, family members and patients tend to rate problems as occurring at the same frequency. On an item by item analysis, there were no statistical differences for ratings on 57 of 70 items. On
the 13 statistically different items, patients rated problems as more frequent than family members. The only scale score that demonstrated statistically different ratings (family vs. patient) was the communication scale ($p<0.01$).

**Advantages.** The reliability of information obtained by self-report from individuals with TBI may be compromised by impaired self-awareness. Incorporating information gathered from collateral sources may compensate for this problem (Hart et al. 2003). The NFI provides for the collection of information from collateral sources allowing for a more comprehensive picture of both the difficulties experienced by the patient and the impact of problems on the home environment (Witol et al. 1999).

**Limitations.** Awad (2002) was unable to establish construct validity for the NFI. The author cited poor fit indices, a large number of items with poor/weak relation to their latent construct (20 items with squared multiple correlations <0.40), strong correlations between subscales and an inability to distinguish a group of individuals with TBI from non-clinical controls as the basis for this assertion. It is suggested that the NFI may be measuring aspects of a single large construct rather than six discrete constructs.

Weingert et al. (1999) reported very low endorsement rates for many of the items resulting in skewed distributions. Low rates of endorsement might indicate that these items are not meaningful discriminators for the head injury population.

While the authors do provide data for comparison, it is not truly normative. The data set used for standardization was derived from a population of individuals with TBI. There is no normative data available based on non-clinical populations (Awad, 2002, Witol et al. 1999).

Although the NFI is widely used, there is relatively little information available in the literature with regard to its reliability, validity and responsiveness. The information that is available pertains to older versions of the NFI and, at present, there are no validity or reliability data available for the 76-item version (Awad, 2002)

**Summary – NFI**

**Interpretability:** Comparative data are provided in the manual stratified by patient age and injury severity. The NFI has been translated into Spanish, German and French.

**Acceptability:** The NFI is a lengthy self-report inventory requiring approximately 20 minutes to complete. Forms are provided for assessment by self or by proxy.

**Feasibility:** The NFI is a proprietary scale and must be purchased.

<table>
<thead>
<tr>
<th>Table 17.41 NFI Evaluation Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td>Rigor</td>
</tr>
</tbody>
</table>
Rancho Los Amigos Levels of Cognitive Functioning Scale (LCFS)

The Rancho Los Amigos Levels of Cognitive Functioning Scale (LCFS) was intended to provide a description of 8 stages of cognitive function through which brain injured patients typically progress during their stay in hospital and acute rehabilitative care (Hagen et al., 1972, Hagen 1982). It was not developed as a scale and is not considered to be an outcome measure. Rather, it is a global index used to describe awareness, environmental interaction and behavioural competence (Timmons et al. 1987, Zafonte et al. 1996). It is used to monitor recovery and classify outcome in patients with brain injury (Gouvier et al. 1987). LCFS rating forms for the original 8-level LCFS are available for download from http://tbims.org/combi. Detailed item descriptions are also available from the website.

Table 17.42 Rancho Los Amigos Levels of Cognitive Functioning

<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
<th>Assistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>No response: Total assistance</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Generalized response: Total assistance</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Localized response: Total assistance</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Confused-agitated: Maximal assistance</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Confused-inappropriate, non-agitated: Maximal assistance</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>Confused-appropriate: Moderate assistance</td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>Automatic-appropriate: Minimal assistance for daily living skills</td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>Purposeful-appropriate: Stand-by assistance</td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>Purposeful-appropriate: Stand-by assistance on request</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Purposeful-appropriate: Modified independent</td>
<td></td>
</tr>
</tbody>
</table>

Hagen et al. 1972. Items in italics are included in the revised Rancho Scale (Hagen, 1997)

Table 17.43 Characteristics of the Rancho Los Amigos Scale

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest: r=0.82 (Gouvier et al. 1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interobserver reliability: average r=0.89 (Gouvier et al. 1987); r=0.84, overall reliability index = 0.91, χ=0.31 (Beauchamp et al., 2001)</td>
</tr>
</tbody>
</table>

| Validity    | Concurrent validity: LCFS ratings correlated with Stover & Zeiger ratings at admission (r=0.92) and discharge from rehabilitation (r=0.73), discharge LCFS ratings also correlated significantly with discharge Glasgow Outcome Scale (GOS) scores (0.76) and expanded GOS scores (0.79) (Gouvier et al. 1987); LCFS ratings and scores on the functional cognition index (FCI) correlated at admission (r=0.79) and discharge (r=0.77) from inpatient rehabilitation (Labi et al. 1998); GCS and LCFS ratings significantly correlated (r=0.329; p<0.05; Hall et al. 1993) |
|-------------| Construct validity (known groups): LCFS ratings could discriminate between groups based on categories of vocational recommendations (return to work, vocational training, supported work and continued remedial therapy; p<0.0001), LCFS ratings accounted for 51% variance between cell means (Mysiw et al. 1989) |
|             | Predictive validity: initial LCFS ratings correlated with Stover & Zeiger ratings (0.65), GOS (r=0.57) and E-GOS (0.73) scores collected at the time of discharge from rehabilitation (Gouvier et al., 1987); LCFS at admission to and discharge from rehabilitation as well as LCFS change scores were significantly associated with |
employment status at one year post-injury (Cifu et al. 1997); initial and discharge LCFS ratings significantly related to vocational status up to 26 months post injury (Rao & Kilgore, 1997)

| Responsiveness | On longitudinal evaluation of treatment medications, LCFS ratings demonstrated significant change (p<0.001) (Rosati et al. 2002); functional improvement in Rancho ratings seen from 3 to 6 months and 6 to 12 months post injury – improvement typically corresponded to improvements in functional performance (Timmons et al., 1987) |
| Tested for ABI/TBI patients? | Specific to brain injury. |
| Other Formats | A revised version incorporates levels of assistance and includes 2 additional levels of Purposeful-appropriate that incorporate varying levels of assistance requirements (Hagen, 1997) |
| Use by proxy? | N/A |

**Advantages.** The LCFS is a quick and simple way to provide a snapshot of an individual’s level of recovery. It is also useful for making quick comparisons between groups (Johnston et al. 1991). Its simplicity and utility have contributed to its widespread use within the United States (Hall and Johnston 1994, Hall 1997).

**Limitations.** At present there is no standardized method to derive an LCFS rating. Variable interobserver agreement has been reported suggesting that standardized rating methods might serve to improve reliability (Beauchamp et al. 2001).

The LCFS provides a quick and simple description of global behaviour from which level of cognitive functioning is inferred. It focuses on the impact of cognitive dysfunction on arousal and overall behaviour, but does not provide information regarding specific domains of cognitive impairment (Labi et al. 1998). There is relatively little published evidence to support the reliability or validity of the LCFS.

**Summary – RLA-CFS**

*Interpretability:* The LCFS is used widely in the United States and provides a quick, global picture of level of recovery.

*Acceptability:* Ratings are derived from observation and represent little or no patient burden. Use of collateral information to derive ratings has not been evaluated.

*Feasibility:* The LCFS is short and simple. It is available free of charge. The LCFS has been evaluated for use in longitudinal assessments.
**Table 17.44 LCFS Evaluation Summary**

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+</td>
<td>+++(TR)</td>
<td>+</td>
</tr>
</tbody>
</table>

*NOTE: +++=Excellent; ++=Adequate; +=Poor; N/A = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.*

### 17.19 Satisfaction with Life Scale (SWLS)

Life satisfaction may be defined as a conscious, cognitive, global judgement of one’s own life. It is not an assessment based on externally imposed objective standards, but rather depends upon a comparison of one’s life circumstances to one’s own internal standards or criteria (Diener et al. 1985, Pavot et al. 1991, Pavot and Diener 1993). The Satisfaction with Life Scale was created to assess a person’s global judgment of life satisfaction (Diener et al. 1985).

Diener et al. (1985) generated 48 self-report items related to satisfaction with life, including items assessing positive and negative affect. Factor analyses were used to identify 3 factors; life satisfaction, negative affect and positive affect. All affect items were eliminated as were items with factor loadings of less than 0.60. The remaining 10 items were reduced to 5 on the basis of “semantic similarity” (Diener et al. 1985).

Respondents are instructed to rate each item using a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Item ratings are summed to provide a total score ranging from 5 – 35 where higher scores are indicative of greater life satisfaction. The SWLS takes a global approach to assessment. Because no specific domains are named within the scale and items are not specific in nature, the respondent remains free to consider the life domains or affective components he or she feels make the most important contribution to their subjective experience of happiness (Diener et al. 1985, Pavot and Diener 1993, Arrindell et al. 1999).

The scale is short and simple to administer and score. It can easily be added to assessments using multiple measures with no significant increase in time (Pavot et al. 1991).

**Table 17.45 Satisfaction with Life Scale* *(Diener et al. 1985)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>In most ways my life is close to my ideal</td>
</tr>
<tr>
<td>2.</td>
<td>The conditions of my life are excellent.</td>
</tr>
<tr>
<td>3.</td>
<td>I am satisfied with my life.</td>
</tr>
<tr>
<td>4.</td>
<td>So far I have gotten the important things I want in life.</td>
</tr>
<tr>
<td>5.</td>
<td>If I could live my life over, I would change almost nothing.</td>
</tr>
</tbody>
</table>
Table 17.46 Characteristics of the Satisfaction with Life Scale

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Test-retest: 0.82 for a 2-month interval (Diener et al. 1985); 0.84 for a 2-week interval and 0.84 for a one-month interval (student sample - Pavot et al. 1991); In their 1993 review, Pavot &amp; Diener (1993) reported test retest reliability ranging 0.83 – 0.50 – intervals ranged from 2 weeks to 4 years and, in general, higher reliabilities were associated with shorter retest intervals. Internal consistency: Item to total correlations ranged from 0.57 – 0.75 (α=0.87) in a sample of undergraduate university students and from 0.63 – 0.81 in a sample of elderly persons (Diener et al. 1985); In a sample of older individuals (mean age = 74), α = 0.83 and 0.85 in a sample of university students (Pavot et al. 1991); Item to total correlations ranged from 0.55 – 0.80 among older individuals and 0.63 – 0.77 among the students; α=0.91(time 1) and 0.82(time 2) – points separated by a few weeks (Suh &amp; Diener 1996); reliability according to Fleishman &amp; Benson formula (1987) was 0.921 (Shevlin et al. 1998); Arrindell et al. (1999) reported α = 0.82 and item-total correlations ranging from 0.5 to 0.7; α=0.78 for the Portuguese version in an adolescent sample (Neto et al. 1993); In a review, Pavot &amp; Diener identified 6 articles evaluating internal consistency – α ranged from 0.79 – 0.89; item-to-total correlations ranged from 0.71 to 0.86—mean interitem correlation = 0.70 α= 0.92 (Westaway et al. 2003); Lucas et al. (1996) reported α = 0.84, 0.84 and 0.88 over three studies; α = 0.78, mean inter item correlation = 0.41, item total correlations ranging from 0.52 – 0.65 (Neto 1993); α=0.86 (Meyer et al. 2004).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
<td>Construct validity: Principal components factor analysis (PCA) revealed a single factor accounting for 66% of the variance – factor loadings ranged from 0.61(Item 5) to 0.84 (Item 1) (Diener et al. 1985); PCA revealed a single factor accounting for 65% and 74% of variance in elderly and student subject samples respectively – loadings ranged from 0.78 – 0.93 (Pavot et al. 1991); PCA revealed a single factor accounting for 60.1% of variance – items 1-4 factor loadings &gt; 70%, item 5 = 0.64 (Arrindell et al. 1999); factor analysis revealed a single factor accounting for 76% of the variance – factor loadings ranged from 0.81 to 0.92 (Westaway et al. 2003); a one-factor measurement model was found for both male and female Spanish adolescents suggesting no factor invariance across the sexes (Atienza et al. 2003); Shevlin et al. (1998) reported a single factor with factor loadings ranging from 0.92 to 0.96; PCA analysis revealed a single factor accounting for 53.3% of variance (Neto 1993); Construct validity (known groups): SWLS scores differentiated between groups of young adults defined by marital status (p&lt;0.001) (Arrindell et al. 1999); significant differences in life satisfaction were identified between all groups of patients based on analyzed disorder (substance use, affective disorder, anxiety disorder, somatoform disorder) and those with no disorder (Meyer et al. 2004). Construct validity (convergent/divergent): SWLS scores correlated with selected personality measures: 0.54 with self-esteem, -0.41 with symptom checklist, -0.48 with neuroticism, -0.25 with emotionality, 0.20 with sociability and very low correlations with activity and impulsivity (Diener et al. 1985); r=0.86 with rated self-esteem (Westaway et al. 2003); r=0.52 (time 1), 0.43 (time 2) with positive affect and r=-0.36 (time 1), -0.30 (time 2) with negative affect (Lucas et al. 1996); r=0.60 (time 1) and 0.52 (time 2) with optimism and r=0.59 (time 1) and 0.55 (time 2) (Lucas et al. 1996); SWLS scores correlated with global happiness (Fordyce Scale r=0.68) as well as with affect balance (r=0.76; Pavot et al. 1991); multi-method multi-trait analyses demonstrated that assessment via the SWLA is able to discriminate between life satisfaction and both affective aspects of SWB, optimism and self-esteem (Lucas et al. 1996); significant positive correlations were</td>
</tr>
</tbody>
</table>
Evidence-Based Review of Moderate to Severe Acquired Brain Injury - 2013

<table>
<thead>
<tr>
<th>Evidence-based Review of Moderate to Severe Acquired Brain Injury</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Updated August 2013</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Section</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrated with social acceptance, self-efficacy, psychological maturity, impulsivity/activity, self-concept, physical attractiveness and happiness while significant negative correlations between SWLS and loneliness, self-assessed loneliness, social anxiety and shyness were reported (Neto 1999); SWLS scores correlated with recent (within 3 months) positive and negative life events ($r=0.25$ and $-0.28$, respectively, $p&lt;0.01$) (Suh et al. 1996).</td>
<td><strong>Concurrent validity:</strong> Moderately strong correlations ($r=0.47-0.68$) with other measures of subjective well-being, including: Fordyce’s % of time happy question and single-item measure of happiness, Differential Personality Questionnaire, Cantril’s Self-Anchoring Ladder, Gurin, Andrews and Withey’s D-T scale, Campbell, Bradburn’s Affect Balance Scale, and Summed Domain Satisfaction – in addition, SWLS scores correlated with interviewer rating of life satisfaction ($r=0.43$) (Diener et al. 1985); Pavot et al. (1991) reported moderate to strong correlations ($r=0.42-0.81$) with both self and peer reported assessments of life satisfaction (LSI-A, Philadelphia Geriatric Morale Scale, Daily satisfaction, memory difference, peer-rated SWLS &amp; peer-rated LSI-A); In a review of studies evaluating SWLS, Pavot &amp; Diener (1993) reported convergence with related measures (Andrews/Withey Scale, Fordyce Global Scale) as well as negative correlations with measures of distress (Beck Depression Inventory, negative affect and anxiety, depression &amp; distress on the Symptom Checklist-90); $r=0.56$ with Oxford Happiness Inventory, $r=0.61$ with the Depression-Happiness scale and neuroticism and conscientiousness were the most significant predictors of SWLS scores (Hayes et al. 2003).</td>
</tr>
<tr>
<td><strong>Responsiveness</strong></td>
<td>From beginning of therapy to one month into the therapy process, SWLS scores changed significantly for clients ($p&lt;0.01$, $n=7$)(Friedman, 1991 in Pavot &amp; Diener 1991); elderly caregivers of patients with dementia demonstrated significant decline in satisfaction with life scores over time ($p&lt;0.05$) (Vitaliano et al.1991).</td>
</tr>
<tr>
<td><strong>Tested for TBI patients?</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Other Formats</strong></td>
<td>The Extended Satisfaction with Life Scale (ESWLS)(Alfonso et al. 1996, Gregg and Salisbury 2001)</td>
</tr>
<tr>
<td></td>
<td>The Temporal Satisfaction with Life Scale (TSWLS) (Pavot et al. 1998, McIntosh 2000)</td>
</tr>
<tr>
<td><strong>Use by proxy?</strong></td>
<td>Pavot et al. (1991) reported correlations between self and peer rated SWLS scores ($r=0.54$) when used to assess elderly individuals (mean age = 74). Among a student population, correlation between peer reports and family reports = 0.54, between self report and peer report = 0.55 and between self and family report = 0.57.</td>
</tr>
</tbody>
</table>

**Advantages.** The scale is available freely and is simple to administer and score. With only 5 items, it takes very little time to complete. The scale has been evaluated for use in populations of varying ages (adolescent, young adult and senior). Scale items appear to be at the 6th – 10th grade reading level, making it comprehensible to a wide range of adults. Original scale was tested in both college student and geriatric populations (Diener et al. 1985). Scale items are at the 6th to 10th grade reading level, which makes it comprehensible to most adults (Pavot and Diener 1993). The scale has been evaluated in several cultures and has been translated into several languages including Dutch, Taiwanese, Spanish, French, Russian, Korean, Hebrew, Mandarin Chinese, Spanish, Spanish and Portuguese.
It has been suggested that social desirability may account for a large part of variance in the assessment of subjective well-being and may, in fact, be an important component of well-being (Pavot and Diener 1993). However, Diener et al. (1985) reported a very weak association between SWLS scores and the Marlowe-Crowne scale of social desirability ($r=0.02$).

**Limitations.** Although the SWLS is used to evaluate satisfaction with life in populations of adults with acquired brain injury, no studies specifically evaluating the use of this scale within this specific population could be identified.

While the SWLS is a simple scale, interpretation of scores is not clear. The SWLS was not intended to provide an assessment of subjective well-being (SWB), only a single aspect of well-being. One cannot assume that SWLS scores provide a direct assessment of emotional well-being. In order to assess the broader construct of SWB, assessment of negative and positive affect should be included (Pavot and Diener 1993). Furthermore, no published normative data for the SWLS could be located. Pavot and Diener (1993) identified numerous studies providing means and standard deviations for SWLS scores in a variety of populations and note considerable variation within different population subsets. However, scores may be interpreted in absolute rather than relative terms. In this case, it has been suggested that a score of 20 is regarded as neutral, while scores in excess of 20 represent satisfaction (21-25=slightly satisfied; 26-30=satisfied), and scores of less than 20 represent dissatisfaction (15-19=slightly dissatisfied; 5-9=extremely dissatisfied) (Pavot & Diener 1993).

The SWLS does not appear to be affected by gender or age (Pavot and Diener 1993). Factor analyses focusing on factorial invariance across gender have demonstrated that the structure and measurement of life satisfaction are equivalent across groups, that is, the strength of relationships between items and the underlying construct is the same for men and women (Wu and Yao 2006, Shevlin et al. 1998). However, factorial invariance was not demonstrated on evaluation of the Spanish version of the SWLS (Atienza et al. 2003, Pons et al. 2000). Westaway et al. (2003) reported that SWLS scores were not related to either gender or age, but rather to employment status and level of education. Similarly, Neto (1993) identified significant main effects associated with both gender and socioeconomic status (SES) such that higher SES and male gender were associated with greater satisfaction with life as assessed on the SWLS.

**Summary – SWLS**

**Interpretability:** Guidelines for absolute interpretation of scores are available. To our knowledge, no normative data is presently available for the SWLS.

**Acceptability:** Scale items are at a suitable reading level for most adults and it takes a minimal amount of time for the subject to complete the measure in its entirety.

**Feasibility:** This scale is brief, simple, and has a low-cost administration.
Table 17.47 SWLS Evaluation Summary

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
<td>++ (TR)</td>
<td>++</td>
</tr>
<tr>
<td>+++</td>
<td>IC</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC = internal consistency; IO = Interobserver.

17.20 Quality of Life after Traumatic Brain Injury (QOLIBRI)

The quality of life after traumatic brain injury (QOLIBRI) questionnaire was designed to specifically measure the quality of life of those who have sustained a TBI (Truelle et al., 2010). Prior to the scale being created the following items were reviewed: Quality of Life of the Traumatic Brain Injured; the Profile de la Qualite de la View Subjective, The Brain Injury Community Rehabilitation Outcome scale and the European Brain Injury Questionnaire. All items from each of the questionnaires were reviewed. Following a review of the items and an assessment of their psychometric properties a preliminary QOLIBRI was developed which consisted of 49 items (von Steinbuchel et al., 2010). These 49 items were arranged into 8 subscales.

The final QOLIBRI consists of 37 items in 6 subscales: cognition (7 items), self (7 items), daily life and autonomy (7 items) and social relationships (6 items), emotions (5 items) and physical problems (5 items). The first 4 subscales are coded on a 1 to 5 scale where 1- not at all satisfied and 5= very satisfied. The responses to the last two subscales (emotion and physical problems) are reverse scored to correspond with the satisfaction items. Here 1- very bothered, and 5= not at all bothered. Responses for each subscale are summed to give a total, then divided by the number of responses to give a scale a mean score. The scale means have a maximum possible range of 1 to 5. The mean can be computed when there are some missing responses, but should not be calculated if more than one third of responses on the scale are missing. In a similar manner the QOLIBRI total score is calculated by summing all the responses, and then dividing by the actual number of responses. Again, a total score should not be calculated if more than one third of responses are missing (http://www.qolibrinet.com/). The scales have also been translated into 7 languages and have been tested with each language cohort.

Table 17.48 The QOLIBRI scale

<table>
<thead>
<tr>
<th>Section</th>
<th>Section</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Your thinking abilities? (self)</td>
<td>Your ability to concentrated (for example when reading or keeping track of a conversation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Your ability to express yourself and understand other in a conversation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Your ability to remember everyday things (for example where you have put things)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Your ability to plan and work out solutions to everyday practical problems (for example what to do when you lose your keys)</td>
</tr>
</tbody>
</table>
## Table 17.49 Characteristics of the QOLIBRI

<table>
<thead>
<tr>
<th>Module 17-Assessment of Outcomes Following Acquired/Traumatic Brain Injury-V9-2013</th>
</tr>
</thead>
</table>
| **Reliability** | **Test-Retest**: Intra class correlations (ICC) from a subsample of 381 participants ranged from 0.78 (emotions) to 0.85 (physical problems) indicating good test-retest reliability. The overall score was 0.91 (CI-0.89 to 0.92) (van Steinbuchel et al., 2010)  
**Internal Consistency**: Internal consistency was assessed for each of the subscales in each language. Cronbach’s $\alpha$ scores ranged from 0.75 (physical problems subscale) to 0.89 (cognition and self subscales). Internal consistency was also found when looking at scores of those with MMSE scores of <28 and comparing them to those with an MMSE score of >27. ($\alpha$ was 0.81 for the physical problems subscale for the group with low cognitive performance, and 0.76 for those with |
normal cognitive status) (von Steinbuchel et al., 2010). Although the QOLIBRI total score is useful as an overall summary, the analysis indicates that it does not completely describe variation in HRQoL and that this is more fully and consistently measured by the profile of individual scales.

### Validity

**Concurrent validity:** There was a significant relationship between the QOLIBRI and the Glasgow Outcome Scale Extended (GOSE). The strongest relationship was with the subscale daily life and autonomy ($r = .42$) and the weakest was with the emotions subscale ($r = .19$). Those with a good recovery reported more areas as good on the HRO0L than those with moderate or severe injuries. The relationship between the HADS and the QOLIBRI was also found to be strong, with the strongest relationship between the HADS depression scale the self scale ($r = -.62$) and the HADS anxiety scale the emotions scale ($r = -.62$). The SF PCS was strongly related to the physical problems scale ($r = .63$) and the SF MCS was strongly associated with the emotions scale ($r = .61$).

**Construct validity:** Further analysis revealed the outcome related information captured by the SF-36 mental health component score was also captured by the QOLIBRI.

**Construct validity (known groups):** Effects of age ($r = .06$), education ($r = .11$), time since injury ($r = -.08$) and the severity of injury as determined by the GCS ($r = -.03$) were all very weak. Current comorbid health conditions showed a significant relationship with all QOLIBRI subscales with the strongest correlation ($r = .56$) on the physical subscale. An association between the test scores and the QOL of the person.

### Responsiveness

N/A

### Tested for ABI/TBI patients?

Developed to be used with those who have sustained an ABI/TBI

### Other Formats

The QOLIBRI was been translated into 7 languages with each scale being found reliable and valid.

### Use by Proxy?

No

### Advantages:

This scale was designed specifically for the ABI population and has been translated into 6 other languages. To date this is the only scale designed specially for those who have sustained either an acquired brain injury or a traumatic brain injury. The composite measure has the advantage of covering both functional outcomes post ABI and HRQoL post ABI.

### Limitations

Like so many other scales measuring quality of life, the important limitation is the complexity of HRQoL, as it remains virtually impossible to capture and define an individual’s view of the future, the concept of individuality and the experience of intimacy (Truelle et al., 2010). The conclusions of the study are based on the approach to recruitment. Here subjects were a convenience sample. Subjects were chosen at various times across a multitude of settings, therefor the sample was scale orientated not patient focused (Truelle et al., 2010).
Summary – QOLIBRI

**Interpretability:** Results are easy to interpret, with lower scores indicating a better QoL.

**Acceptability:** The scale, available in 7 languages, is a self-report of how well the individual perceives themselves as doing.

**Feasibility:** The scale is now available and ready for more regular use. It is easy to use, available in a variety of languages and there is no fee for its use.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigor</td>
<td>Results</td>
<td>Rigor</td>
</tr>
<tr>
<td>+++</td>
<td>+++ (TR)</td>
<td>+++</td>
</tr>
<tr>
<td>++</td>
<td>++ (IC)</td>
<td>-</td>
</tr>
</tbody>
</table>

**NOTE:** +++=Excellent; ++=Adequate; +=Poor; n/a = insufficient information; TR=Test re-test; IC= internal consistency; IO = Interobserver.
17.21 Reference List


Malec J.F., Lezak M.D. Manual for The Mayo-Portland Adaptability Inventory (MPAI-4) 1-77.


