Reliability and Validity of the Face, Legs, Activity, Cry, Consolability Behavioral Tool in Assessing Acute Pain in Critically Ill Patients

By Terri Voepel-Lewis, RN, MSN, Jennifer Zanotti, RN, MS, CCRN, CEN, Jennifer A. Dammeyer, RN, MSN, and Sandra Merkel, RN, MS

Background
Few investigators have evaluated pain assessment tools in the critical care setting.

Objective
To evaluate the reliability and validity of the Face, Legs, Activity, Cry, Consolability (FLACC) Behavioral Scale in assessing pain in critically ill adults and children unable to self-report pain.

Methods
Three nurses simultaneously, but independently, observed and scored pain behaviors twice in 29 critically ill adults and 8 children: before administration of an analgesic or during a painful procedure, and 15 to 30 minutes after the administration or procedure. Two nurses used the FLACC scale, the third used either the Checklist of Nonverbal Pain Indicators (for adults) or the COMFORT scale (for children).

Results
For 73 observations, FLACC scores correlated highly with the other 2 scores (ρ = 0.963 and 0.849, respectively), supporting criterion validity. Significant decreases in FLACC scores after analgesia (or at rest) supported construct validity of the tool (mean, 5.27; SD, 2.3 vs mean, 0.52; SD, 1.1; P < .001). Exact agreement and κ statistics, as well as intraclass correlation coefficients (0.67-0.95), support excellent interrater reliability of the tool. Internal consistency was excellent; the Cronbach α was 0.882 when all items were included.

Conclusions
Although similar in content to other behavioral pain scales, the FLACC can be used across populations of patients and settings, and the scores are comparable to those of the commonly used 0-to-10 number rating scale. (American Journal of Critical Care. 2010;19:55-62)
Frequent and routine assessment of pain improves pain management for adults and children and is considered essential for optimal care. Additionally, clinical practice guidelines for the use of sedatives and analgesics in critically ill patients highlight the importance of systematically and consistently assessing and documenting pain and response to therapy by using scales appropriate for the population of patients. These guidelines, as well as previous reports, suggest that pain assessment for patients who cannot communicate their pain should include subjective observation of pain-related behaviors (eg, movement, facial expression, posturing). Despite such recommendations and pain standards from the Joint Commission, considerable gaps exist in pain assessment practices in critical care because of the limited research in this area.

Several investigators have generated similar, qualitative descriptors of pain behaviors in adults and children with cognitive impairment and in critically ill adults and children. For instance, Mateo and Krenzischek reported moderate correlations between the degree of facial grimacing, muscle tension, and sounds documented by a nurse and the verbal description of pain reported by patients in the postanesthesia care unit. In another study, Puntillo et al compared nurses’ subjective ratings of pain, number of behavioral indicators (eg, movements, facial expression, posturing), physiological parameters, and patients’ ratings in 31 critically ill surgical patients and found moderate correlations between nurses’ ratings and number of behavior indicators, and between nurses’ and patients’ ratings.

Such data have led to the development of behavioral scales, including simple scales such as the Checklist of Nonverbal Pain Indicators (CNPI), the Behavioral Pain Scale (BPS), and the Critical-Care Pain Observation Tool (CPOT). Almost all behavioral pain scales require some grading or scoring of facial expression, vocalizations, and bodily movements. The CNPI requires simple scoring of each of 6 behaviors (vocalizations, grimaces, bracing, rubbing, restlessness, verbal complaint) as present or absent, to provide a total score of 0 to 6. The BPS requires grading of 3 categories (facial expression, upper limb movement, and compliance with ventilation) to provide a score of 3 to 12. The CPOT requires grading each of 4 behavioral categories (facial expression, body movements, muscle tension, and vocalization or compliance with ventilator) on a scale of 0 to 2 to provide a total score of 0 to 8. The COMFORT scale, which has been widely studied in children, contains 8 categories (alertness, calmness, respiratory response, physical movement, muscle tone, facial tension, heart rate, and blood pressure); each category is scored from 1 to 5 to produce a total score of 8 to 40.

Each of these tools has good interrater agreement and good validity in differentiating noxious stimuli (eg, turning) from rest or pain-free situations.

About the Authors

Terri Voepel-Lewis is a research area specialist and Sandra Merkel is a clinical nurse specialist, Department of Anesthesiology, and Jennifer A. Dammeyer is a clinical nurse specialist, Department of Critical Care, in the University of Michigan Health System, Ann Arbor, Michigan. Jennifer Zanotti is a clinical nurse specialist, Emergency Services, at Memorial Health System, Colorado Springs, Colorado.

Corresponding author: Terri Voepel-Lewis, RN, MSN, Department of Anesthesiology, Section of Pediatrics, F3900 C.S. Mott Hospital, SPC 5211, 1500 E Medical Center Dr, Ann Arbor, MI 48109-5211 (e-mail: terriv@umich.edu).
These studies indicate that observing behaviors and using simple scales can be effective in assessing pain in nonverbal patients.

To be clinically useful, pain assessment tools must be readily adaptable in busy settings such as the intensive care unit. Several characteristics affect the clinical usefulness of an assessment tool, including the tool’s relative advantage compared with other tools, its compatibility (how similar the instrument is to other tools already used in the setting), and its complexity (ease of use). Furthermore, the ability to use a single tool in different populations of patients may improve the clinical usefulness of the tool.

Many observational pain scales lack these qualities. For instance, the most commonly used and recommended verbal self-report tool is the 0-to-10 number rating scale (NRS), in which 0 indicates no pain and 10 indicates worst pain. Many observational tools, including those developed for critical care, have scales that differ from the 0-to-10 format, potentially confusing the clinical interpretation of pain scores. In contrast, with the FLACC tool, each of 5 behavioral categories, facial expression, leg movement, bodily activity, cry or verbalization, and consolability, is rated on a scale of 0 to 2 to provide an overall pain score ranging from 0 to 10, consistent with the NRS.

The FLACC Behavioral Scale includes behavioral categories and a variety of descriptors that are reliably associated with pain in children, adults with cognitive impairment, and critically ill adults, supporting the content validity of the tool in these populations. The FLACC tool is widely recognized and used in the United States and internationally and has been translated into several languages, including French, Chinese, Portuguese, Swedish, and Italian. Last, the tool in a revised form has a high degree of clinical usefulness in assessing pain in children with cognitive impairment, and critically ill adults, who were present in any of the critical care units in the medical center during the study period (2002-2004). Patients were included if they could not self-report their pain (eg, because of intubation with or without change in cognition), and if they had an underlying condition associated with pain or were undergoing a procedure known to cause pain. Patients receiving muscle relaxants were excluded.

**Data Collection**

Observations were made by 3 intensive care unit nurses during the routine care of each patient as follows: Before administration of an analgesic, or during a painful procedure such as turning or suctioning, nurses observed the patient and simultaneously, but independently, scored pain behaviors during a 1- to 2-minute period. Nurses had no knowledge of the scores of their fellow nurses. Two of the nurses used the FLACC tool to score pain behaviors; the third nurse used the CNPI for adults and the COMFORT Scale for children. Each patient was observed again by the same nurses approximately 15 to 30 minutes after the first observation. Patients’ demographics, illness, type of procedure, and analgesic administered were recorded.

### Table 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td></td>
</tr>
<tr>
<td>No particular expression or smile</td>
<td>Occasional grimace, frown, withdrawn or disinterested</td>
</tr>
<tr>
<td>Legs</td>
<td></td>
</tr>
<tr>
<td>Normal position or relaxed</td>
<td>Uneasy, restless, or tense</td>
</tr>
<tr>
<td>Activity</td>
<td></td>
</tr>
<tr>
<td>Lying quietly, normal position, moves easily</td>
<td>Squirming, shifting back and forth, or tense</td>
</tr>
<tr>
<td>Cry</td>
<td></td>
</tr>
<tr>
<td>No cry</td>
<td>Moans, whimpered, or occasional complaint</td>
</tr>
<tr>
<td>Consolability</td>
<td></td>
</tr>
<tr>
<td>Content, relaxed</td>
<td>Reassured by occasional touching, hugging, or being talked to; distractible</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Printed with permission © The Regents of the University of Michigan.*
Data Analyses

SPSS software (SPSS Inc, Chicago, Illinois) was used to analyze the data. Total FLACC and CNPI scores were treated as ordinal data, and each category within the FLACC was treated as ordinal, poly-chotomous data, as recommended and used by previous investigators.27-29 Interrater reliability was evaluated by using intraclass correlation coefficients, which determine the strength of association and measure of chance-corrected agreement. Additionally, exact agreement for scores within each of the 5 FLACC categories was evaluated by using κ statistics. In accordance with well-established criteria,30 interrater agreement for total FLACC scores was considered excellent at an intraclass correlation coefficient of 0.75. Because each FLACC category contains only 3 items, generating comparatively less variance and thereby limiting the magnitude of correlations,31 an intraclass correlation coefficient of 0.41 was accepted as adequate agreement, and a coefficient of 0.6 was considered good to excellent agreement.32

Criterion Validity

FLACC scores correlated significantly with CNPI scores, supporting excellent criterion validity in adults (ρ = 0.963; P < .01). Additionally, FLACC and COMFORT scores were highly correlated (ρ = 0.849; P < .01), supporting criterion validity in critically ill children.

Construct Validity

FLACC pain scores decreased significantly after administration of an analgesic or from painful to nonpainful situations (mean, 5.27; SD, 2.3 vs mean, 0.52; SD, 1.1; P < .001), supporting excellent construct validity across populations of patients.

Reliability

Agreement was excellent between observers for each category of the FLACC, as well as for total FLACC scores, supporting the interrater reliability of the tool for paired data to compare scores before and after analgesic administration or during and after a painful procedure. P values less than .05 were accepted as significant. The internal consistency of reliability of the items in the FLACC tool was measured by using Cronbach (coefficient) α values. Cronbach α values of 0.7 or greater were considered indicative of excellent internal consistency. A principal component and exploratory factor analysis were performed to identify underlying factors that explained the variance in the FLACC total scores; loading factors of 0.45 or greater were considered acceptable.

Sample Size

The sample size was conservatively based on a moderate reliability correlation coefficient between FLACC scores. For α = 0.05 and β = 0.1, a total of 25 observations would be needed to reveal a modest correlation of at least 0.6.33 A minimum of 65 observations with at least 13 paired observations (eg, before and after analgesia) would be needed to ensure a sufficient number of FLACC scores across the spectrum (ie, mild, moderate, and severe pain scores). This sample size would be sufficiently large to satisfy the stronger correlations required for criterion validity (ie, r = 0.75) and to establish a minimum decrease in pain scores from a mean of 5.3 (SD, 2.8) to a mean of 2 (SD, 2.4).

Results

A total of 73 observations were obtained in 29 critically ill adults and 8 children. Table 2 gives a description of the patients.

Table 2: Description of the sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Adults (n = 29)</th>
<th>Children (n = 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>51 (38)</td>
<td>5.6 (5.6)</td>
</tr>
<tr>
<td>Range</td>
<td>24 - 70</td>
<td>0.13 - 13</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17 (59)</td>
<td>4 (50)</td>
</tr>
<tr>
<td>Female</td>
<td>12 (41)</td>
<td>3 (38)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>23 (79)</td>
<td>8 (100)</td>
</tr>
<tr>
<td>African American</td>
<td>3 (10)</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>3 (10)</td>
<td>0</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>19 (66)</td>
<td>1 (12)</td>
</tr>
<tr>
<td>Neurological-medical</td>
<td>6 (21)</td>
<td>0</td>
</tr>
<tr>
<td>Neurosurgical</td>
<td>2 (7)</td>
<td>0</td>
</tr>
<tr>
<td>Surgical</td>
<td>2 (7)</td>
<td>7 (88)</td>
</tr>
<tr>
<td>Treated with mechanical ventilation</td>
<td>23 (79)</td>
<td>3 (38)</td>
</tr>
<tr>
<td>Cognitive impairment</td>
<td>16 (55)</td>
<td>1 (12)</td>
</tr>
<tr>
<td>Acute delirium</td>
<td>11 (38)</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

a Values are number (%) unless indicated otherwise. Because of rounding, not all percentages total 100.

b Data on sex missing for 1 child.
in assessing pain in critically ill patients (Table 3). Agreement was also adequate to excellent when data on adults, children, and patients receiving mechanical ventilation were analyzed separately (Table 3).

### Internal Consistency and Factor Analysis

Internal consistency of the FLACC was excellent, as indicated by Cronbach $\alpha = 0.882$, when all items were included. Each category correlated highly with the others (Spearman $\rho = 0.69-0.92$; $P < .001$) except for the cry category ($\rho = 0.18-0.36$). Additionally, the Cronbach $\alpha$ improved to 0.934 when the cry category was removed, but decreased slightly with removal of other items. In the exploratory factor analysis, 1 component accounted for 68.9% of the variance in FLACC scores; 4 items contributed to this component: face (0.86), legs (0.94), activity (0.90), and consolability (0.95). These findings indicate that 4 categories of the FLACC reflected the pain expression factor in this sample of patients.

### Discussion

Use of behavioral pain tools may help in assessing pain in critical care patients, but the tools must have good reliability and validity and be clinically feasible. Clinical feasibility, or the ability to readily adapt an instrument for routine assessment and documentation, may depend on a tool’s simplicity and its compatibility with other tools used in the clinical setting, as well as on the ability to use the tool across settings or populations of patients. We evaluated the well-known FLACC behavioral pain tool and showed that the tool has excellent interrater reliability, criterion validity, and construct validity, thereby supporting its usefulness in assessing pain in critical care patients.

Indisputably, self-report remains the gold standard for pain assessment, yet many patients cannot report their pain, an inability that may make them vulnerable to poor pain management. Many tools have been developed to aid in assessing pain for patients who cannot self-report; however, few of the tools have been tested in critically ill patients who cannot self-report. We found that the FLACC Behavioral Scale has excellent psychometric properties, including reliability, criterion validity, and construct validity, in assessing pain in these patients. Interestingly, 4 categories (face, legs, activity, and consolability) were predictive of most of the variance (68.5%) in scores. The cry category correlated poorly with other categories and slightly lowered the internal consistency of the tool. These findings are not surprising; many of the patients in our study were nonverbal and many had endotracheal tubes.

The COMFORT Scale, BPS, and CPOT, which were all developed for scoring pain in the intensive care unit, include a category for assessing respiratory response or compliance with ventilation, a category that may be useful for assessing pain in patients receiving mechanical ventilation. In a sample of sedated adults receiving mechanical ventilation, compliance with mechanical ventilation had a smaller,
Behavioral pain tools assess the patient’s expressions of distress and discomfort. Behavioral pain scores must be interpreted in light of the patient’s medical condition, including response to analgesia.

but significant, coefficient in accounting for variance in pain expressions, supporting the inclusion of compliance descriptors in tools used to assess pain in patients receiving mechanical ventilation. However, a recent study validating use of the BPS in sedated patients suggested that newer modes of ventilation that allow for variation in patients’ needs may reduce the reliability of this category in assessing discomfort. Interestingly, in our study, the FLACC had good reliability in assessing pain even in the subset of patients receiving mechanical ventilation. However, the addition of descriptors (e.g., breath holding, splinting, blocking ventilation) in the cry category that allow for scoring pain in patients who are intubated and receiving mechanical ventilation may enhance pain assessment in these patients. Indeed, similar minor revisions related to respiratory patterns, in addition to other revisions, improved the reliability of the FLACC tool in assessing pain in cognitively impaired children.

Several guidelines suggest that in addition to observation of behaviors, pain assessment in the critically ill should include consideration of physiological measures such as heart rate, blood pressure, and respiratory rate. Importantly, changes in these measures are nonspecific to pain and may indicate other pathological changes. In a recent study of the COMFORT scale in the pediatric intensive care unit, 97% of the variance in pain scores was explained by 6 behavioral categories, including a category for scoring respiratory or compliance behaviors, but not by heart rate or blood pressure. These findings led the authors to conclude that these parameters should be removed from the COMFORT scale.

The fact that behavioral pain tools provide a score of a patient’s expressions of distress and discomfort must be emphasized. In addition to pain, these behaviors have many potential underlying sources, including physiological abnormalities (e.g., cardiorespiratory compromise) and anxiety. Such conditions are common in critically ill patients, and therefore a patient’s medical condition and current circumstances, including response to analgesia, must be considered when behavioral pain scores are interpreted.

Additionally, most behavioral pain tools, including the FLACC, COMFORT, BPS, and CPOT, were developed to score intensity of acute pain. It has been suggested that behavioral distress related to pain lessens over time, despite persistence of pain. Withdrawn or disinterested expressions and immobility may replace behaviors such as grimacing, vocalizations, and movements. The variety of descriptors included in the FLACC tool were meant to indicate some of the differences observed from patient to patient. However, assessment of chronic or long-term pain should include other observations such as activity, quality of sleep, and expressions of depression.

The ability to generalize our findings may be limited by the following design issues. First, the same nurses scored pain before and after administration of analgesics, a practice that could have resulted in a reporting bias. However, in previous studies in which nurses were blinded to treatment, similar changes in FLACC scores occurred, providing some external validity to our data. Second, we included a variety of medical and surgical patients in the sample to indicate usefulness across critical care settings. However, because of the small sample size, we could not analyze data separately for each group. Further study in these subsets of patients may provide greater insight into behavioral changes that best describe pain in these groups.

## Conclusion

The FLACC behavioral pain tool has excellent reliability and validity in assessing pain in critically ill adults and children. Although similar in content to other observational pain scales, the FLACC tool may offer an advantage: it can be used across populations and settings, and FLACC scores are comparable to scores generated by using 0-to-10 number rating scales.

## Acknowledgments

This study was conducted at the University of Michigan Health System in Ann Arbor.

## Financial Disclosures

None reported.

## eLetters

Now that you’ve read the article, create or contribute to an online discussion on this topic. Visit www.ajcconline.org and click “Respond to This Article” in either the full-text or PDF view of the article.

## See Also

For more about pain assessment, visit the Critical Care Nurse Web site, www.ccnonline.org, and read the article by Kabes et al, “Further Validation of the Nonverbal Pain Scale in Intensive Care Patients” (February 2009).
1. Which of the following patient types have been studied previously for validation of the Face, Legs, Activity, Cry, Consolability (FLACC) Behavioral Scale?
   a. Acutely ill children
   b. Adults with dementia and cognitive impairment
   c. Critically ill adults and children
   d. Adults with expressive aphasia

2. Which of the following is most likely to cause a critically ill patient to be unable to self-report his or her level of pain?
   a. Inability to communicate
   b. Presence of an endotracheal tube
   c. Inadequate nutrition
   d. Sleep deprivation

3. Which of the following is cited by the authors to explain why gaps exist in pain assessment practices in critical care?
   a. Recommendations of the Joint Commission are not followed.
   b. Pain assessment is difficult when the patient is sedated.
   c. Limited research has been done on pain assessment practices in critical care.
   d. Multiple scales for pain assessment are available to critical care nurses.

4. Which of the following are requirements common to almost all behavioral pain scales?
   a. Some grading or scoring of respiratory rate, heart rate, and blood pressure
   b. Some grading or scoring of facial expressions, vocalizations, and body movements
   c. Some grading or scoring of cognition, level of consciousness, and attention span
   d. Some grading or scoring of response to painful stimulation by posturing or withdrawal

5. Which of the following are the categories for assessment in the FLACC pain tool?
   a. Face, activity, compliance with the ventilator, cry
   b. Face, limbs, activity, sounds, calmness
   c. Face, legs, physical movement, cry, calmness
   d. Face, legs, body activity, cry, consolability

6. Which of the following characteristic may affect the clinical usefulness of a pain assessment tool?
   a. Complexity (ease of use)
   b. Unique (uses different factors for evaluation)
   c. Electronic (computerized documentation)
   d. Reading level (6th grade)

7. How is pain assessed with the Number Rating Scale?
   a. Verbal self-report using common descriptive terms for pain
   b. Five behavioral categories that are each rated on a scale of 0 to 2
   c. Verbal self-report where 0 indicates no pain and 10 indicates worst pain
   d. Pain is reported by the nurse on a scale of 0 to 10 based on patient assessment findings

8. Which of the following patient types were excluded from the study?
   a. Intubated patients
   b. Patients undergoing a procedure known to cause pain
   c. Patients who could not self-report pain
   d. Patients receiving muscle relaxants

9. What is considered the "gold standard" for pain assessment?
   a. Self-report
   b. Use of a behavioral pain assessment tool
   c. Wong-Baker faces
   d. Use of the Numeric Rating Scale

10. Which of the following behavioral observations would result in a FLACC score of 0?
    a. No expression, relaxed position, lying quietly, no cry, and appears content
    b. Grimace, restless, moves easily, no cry, and can be distracted
    c. Clenched jaw, kicking, rigid movement, moans, and can be reassured
    d. No expression, restless, moves easily, no cry, and can be reassured

11. What conclusion did the investigators reach related to the use of the FLACC behavioral scale in assessing pain?
    a. The FLACC tool has validity in assessing pain in specific populations.
    b. The FLACC tool has no reliability and validity in assessing pain in critically ill adults.
    c. The FLACC tool had excellent reliability and validity in assessing pain in critically ill adults.
    d. The FLACC tool has poor correlation to scores generated by using 0 to 10 number rating scales.

The American Association of Critical-Care Nurses is accredited as a provider of continuing nursing education by the American Nurses Credentialing Center’s Commission on Accreditation. AACN has been approved as a provider of continuing education in nursing by the State Boards of Nursing of Alabama (#ABNP0062), California (#H1036), and Louisiana (#ABN12). AACN programming meets the standards for most other states requiring mandatory continuing education credit for relicensure.
Reliability and Validity of the Face, Legs, Activity, Cry, Consolability Behavioral Tool in Assessing Acute Pain in Critically Ill Patients
Terri Voepel-Lewis, Jennifer Zanotti, Jennifer A. Dammeyer and Sandra Merkel

Am J Crit Care 2010;19 55-61 10.4037/ajcc2010624
©2010 American Association of Critical-Care Nurses
Published online http://ajcc.aacnjournals.org/

Personal use only. For copyright permission information:
http://ajcc.aacnjournals.org/cgi/external_ref?link_type=PERMISSIONDIRECT

Subscription Information
http://ajcc.aacnjournals.org/subscriptions/

Information for authors
http://ajcc.aacnjournals.org/misc/ifora.xhtml

Submit a manuscript
http://www.editorialmanager.com/ajcc

Email alerts
http://ajcc.aacnjournals.org/subscriptions/etoc.xhtml