Pain assessment and management are components of the Joint Commission 2001 standards of care that require every patient to be assessed for pain. Assessment and management of pain are essential to provision of quality care in all settings. Several valid and useful scales for rating the intensity of pain in most children and adults are available. These various verbal (no pain to worst pain), pictorial (Faces scale), and numeric (1-10) scales for rating pain are commonly used with alert adult or child patients in hospitals and home care. These scales require patients to be able to cognitively indicate a position on a line from no pain to worst possible pain, select a picture that expresses their pain level, or select a number between 1 and 10 to represent their pain level. It is important to differentiate between pain intensity ratings based on patients’ self-reports of pain severity and pain behavior scales that list a number of observable characteristics that can only indicate the presence or apparent absence of pain in patients who cannot self-report.

Observational rating scales, such as the Faces, Legs, Activity, Cry, and Consolability scale (FLACC), for infants and preverbal children are also commonly used. Recently, interest has increased in the development and validation of observational pain scales for use in cognitively impaired patients and critically ill patients who are sedated and receiving mechanical ventilation in intensive care units (ICUs). A group of Canadian nurse investigators described observable physiological and behavioral indicators of pain in critical care patients receiving mechanical ventilation and noted that pain documentation was incomplete or inadequate, adding that lack of a pain assessment tool was most likely a contributing factor. Observational indicators of behaviors associated with pain can vary by patient populations. Pasero and McCaffery described how ratings based on ventilator compliance should not be appropriate in patients who are not receiving mechanical ventilation and pointed out that heavily sedated patients may have severe pain but be unable to move. Such patients will have low scores on observational ratings that use...
ventilator compliance or leg or arm movements as indicators. For ICU patients who cannot self-report, observational behavioral rating scales have been developed for rating pain. Only limited reports of the validity and reliability of those scales have been published.

In a search for an acceptable pain assessment tool for these patients, staff at Creighton University Medical Center Hospital initially tried the FLACC (FLACC: Behavioral Pain Scale in the ICU. The FLACC scale was designed to rate indicators of pain in infants and preschool children and is used extensively for that purpose. Wegman4 noted that the FLACC scale has been used in adult patients as an observational pain scale, primarily because of the Joint Commission requirement for pain assessment of all patients and the difficulty in finding valid and reliable nonverbal pain scales appropriate for adults who have cognitive impairments or are sedated and receiving mechanical ventilation. The ICU nurses considered the FLACC scale unsatisfactory for critically ill adult patients because it includes crying behaviors and reactions to comforting methods, which vary greatly in adults. Thus, the nurses did not use the FLACC consistently.

A subsequent review of the literature located 2 scales designed for use in adult patients in the ICU who are sedated, comatose, and/or receiving mechanical ventilation; the Behavioral Pain Scale (BPS) and the Nonverbal Pain Scale (NVPS). These scales were of particular interest because they addressed these particular ICU problems. The hospital research council reviewed the literature on both scales and selected the NVPS for further testing in the ICU. This selection was based on the preference of these clinicians for a tool that included physiological indicators. The BPS focuses on behavioral observations only (facial expression, cry, and movements), whereas the NVPS includes behavioral and physiological indicators. Although the ICU nurses recognized that physiological indicators should not be used as the sole indicators of pain level, they thought that a combination of behavioral and physiological indicators of pain would be more comprehensive than use of behavioral indicators alone. Additionally, they thought that use of the BPS might be confusing because the maximum score is 8 and that might be seen as lesser pain when compared with the maximum score of 10 used in other pain scales.

Two versions of the NVPS have been described. The original NVPS rated facial expression, activity, guarding, change in vital signs (physiologic I), and other physiological signs (physiologic II; see rating levels for each item in Table 1). Odhner et al6 reported an acceptable level of internal consistency (α = 0.74) and interrater reliability (0.78) for this initial version, but noted that the physiologic II category discriminated less well than did the other subscales.

Wegman4 reported a revision in the NVPS in which a respiratory category was substituted for the former physiologic II category. The new respiratory category included ratings of baseline respiratory rate, oxygen saturation as measured by pulse oximetry, and compliance with the ventilator (see Table 1, item 5A for rating levels). This revision was described in a presentation and a letter to the editor,4 but no further testing was reported.

Because the original NVPS had been validated only once by the authors,6 and the validation of the revised version had not been published, the research council made a decision to test both the initial and revised items of the NVPS to compare the validity and reliability of the 2 versions. The purpose of this study was to further validate a nonverbal pain scale for ICU patients who are sedated, receiving mechanical ventilation, or otherwise unable to express their pain. The research question was as follows: Which version of the NVPS is the most valid and reliable in sedated ICU patients receiving mechanical ventilation?

**Methods**

The study was nonexperimental and methodological. The study was
given exempt status from the institutional review board at Creighton University Medical Center Hospital and Creighton University School of Nursing because all the patients were incapable of self-reporting their pain or providing informed consent for participation. The patients faced no additional risk because this research involved traditional nursing care procedures such as suctioning or repositioning a patient and the research was gathered by using observational data only. Each patient’s privacy and confidentiality were protected by assigning a case number to each patient’s data.

**Sample**

The sample was a convenience sample of patients in the ICU at Creighton University Medical Center Hospital. The ICU is a 25-bed unit that primarily serves trauma and surgical patients. Subjects were at least 19 years old and were unable to verbalize or otherwise indicate pain by using a traditional scale. Patients excluded from the study were those receiving paralytic medications such as cisatracurium besylate or vecuronium bromide and those patients who were paralyzed without the use of medications or had been declared brain dead.

**Procedures**

Data collectors were individually trained by the principal investigator (A.M.K.), who provided raters with a PowerPoint presentation that included criteria for selecting appropriate patients for the study and the data collection tool. The presentation contained the inclusion and exclusion criteria. During the training session, each rater practiced using the tool with the principal investigator. Once each nurse was comfortable and had achieved 90% interrater agreement with the principal investigator, the nurse was allowed to participate in the study.

Table 1 Data collection tool from the study of the Nonverbal Pain Scale#

<table>
<thead>
<tr>
<th>Categories</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Pre-score</th>
<th>Intervention Score</th>
<th>Post Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Face</td>
<td>No particular expression or smile</td>
<td>Occasional grimace, tearing, frown or wrinkled forehead</td>
<td>Frequent grimace, tearing, frown or wrinkled forehead</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Activity (movement)</td>
<td>Lying quietly, normal position</td>
<td>Seeking attention through movement of slow cautious movements</td>
<td>Restless activity and/or withdrawal reflexes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Guarding</td>
<td>Lying quietly, no positioning of hands over areas of body</td>
<td>Splinting areas of the body, tense</td>
<td>Rigid, stiff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Physiologic I (vital signs)</td>
<td>Stable vital signs, no change in past 4 hours</td>
<td>Change over past 4 hours in any of the following: SBP &gt;20 HR &gt;20 RR &gt;10</td>
<td>Change over past 4 hours in any of the following: SBP &gt;30 HR &gt;25 RR &gt;20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5A. Respiratory</td>
<td>Baseline RR/SpO₂ Complaint with ventilator</td>
<td>RR &gt;10 above baseline or 5% ↓ SpO₂ Mild asynchrony with ventilator</td>
<td>RR &gt;20 above baseline or 10% ↓ SpO₂ Severe asynchrony with ventilator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5B. Physiologic II</td>
<td>Warm, dry skin</td>
<td>Dilated pupils, perspiring, flushing</td>
<td>Diaphoretic, pallor</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**5A TOTAL**

**5B TOTAL**

Abbreviations: HR, heart rate; RR, respiratory rate; SBP, systolic blood pressure; SpO₂, oxygen saturation as measured by pulse oximetry.

# Reprinted with permission of Strong Memorial Hospital, University of Rochester Medical Center, developer and copyright holder of the scale.
Data were entered by an administrative assistant who had been trained about the Health Insurance Portability and Accountability Act.

The data were collected by 9 registered nurses, 2 men and 7 women, in the ICU. Among the nurses, 2 had associate’s degrees, 3 had nursing diplomas, 3 had bachelor’s degrees, and 1 had completed graduate level classes. Three of these nurses were charge nurses. Of the 9 nurses, 3 worked the night shift, and 6 worked the day shift. The nurses had from 1 year to more than 20 years of experience working in an ICU. Two were men, and 7 were women. A total of 87% of the data (105 of 121 paired observations) was collected by the principal investigator, and the 3 charge nurses together accounted for 67% of the data collected (81 of the 121 paired observations). All patients were assessed 3 times: before, during, and at rest after a painful nursing procedure. For each patient, initial assessment at rest was immediately before the painful stimulus, and further observations were made during the procedure. Follow-up assessment was done 2 to 10 minutes after the painful stimulus.

A data collection tool was created that combined the original NVPS’ items with the revised item from the NVPS. This tool (Table 1) included the 4 common items: activity/movement, facial expression, guarding, and physiologic I (vital signs). It also included both the original physiologic II item and the revised (respiratory/ventilator compliance) item. This combination provided a simple 6-item data collection tool that could be analyzed by comparing the contributions of the 5 pain items of the original with the added revised item (Table 1). This combined tool was chosen over using 2 distinct forms of the tool because the combined form was simpler for the raters and less likely to lead to reluctance of the busy ICU nurses to participate in the study. Because only a single item was changed in the revised instrument, it was deemed redundant to collect data on the other 4 identical items twice. This method was consistent with traditional methods of scaling and ranking based on the amount of a construct (pain) with an additional item that rates pain in the same manner, from lower to higher on a 3-point ordinal scale. This adjustment is equivalent to a monotonic scale calibration and should not affect results. Statistical consultation supported the notion that researchers use many related items in validating a scale by ensuring that the number of subjects per item is adequate and subsequently eliminating any inadequate item or items. This process is called item analysis or reliability analysis in classical test theory. Special software (which was not available) is needed to test the complex correlation, so the correlation between the original and the modified versions was calculated, and SPSS (SPSS Inc, Chicago, Illinois) software was used to detect any significant positive correlation. A total of 121 independent observations by pairs of nurses on 64 different patients in the ICU were used for data analysis. For observations, pairs of nurses independently rated the same patient. Some patients were observed and rated more than once.

**Data Analysis**

Interrater reliability was established by percent agreement procedures. For each set of concurrent observations, the difference between raters on the total score for the original NVPS and for the revised NVPS was calculated. When the difference was 0 or 1, the percent agreement was determined to be within 90%.

Internal consistency of both the original and revised NVPS was measured to determine if the 5 items on each scale and the total scale were consistently measuring discomfort in patients. In addition, the correlation of each item to the relevant total scale score was computed. The Cronbach $\alpha$ was calculated for the data collected before, during, and after the painful stimulus. Item to total correlations were also used to explore internal consistency. Spearman correlation was used because of the skewed distribution of the data (Figure 1).

Hypothesis testing of relationships is an accepted approach to testing construct validity. The theoretical construct, NVPS pain observations, was tested by hypothesizing predicted relationships between pain ratings of patients at rest and during a discomfort-inducing procedure (suctioning or repositioning). The hypothesis tested was that observations of pain ratings at rest would be significantly lower before a distressing procedure than during the procedure and would decrease approximately 2 to 10 minutes after the procedure.

Construct validity was assessed by using the Friedman repeated-measures test (the nonparametric analogue of repeated-measures analysis of variance) to test whether
the hypothesized differences in ratings occurred among the 3 times (before, during, and after the intervention). An advantage of a repeated-measure (also called dependent or correlated measures) design is that each person serves as his or her own control, decreasing error variance and yielding a more powerful test with the specific sample size. Because of the skewed nature of the data, analysis of variance with repeated measures would not have been appropriate. The Wilcoxon signed rank test was used for post hoc testing to examine where (between which ratings) the differences were significant. Table 2 is a glossary of statistical terms with source citations used in this section. All analyses were done separately on the original and revised versions of the NVPS.

**Results**

Interrater reliability assessments met the 90% agreement criterion in most of the comparisons. On a subset of 76 concurrent observations by 2 nurses, agreement of at least 90% was achieved on 72 observations (94.7%) with the original NVPS and on 69 observations (90.8%) with the revised NVPS.

Testing of the original scale resulted in internal consistency as indicated by Cronbach α values of 0.36 (prior), 0.62 (during), and 0.62 (after). The revised scale resulted in α values of 0.36 (prior), 0.72 (during), and 0.71 (after). Spearman item to total correlations (Table 3) showed moderate correlations for most of the items with their respective scale. Correlations between the physiologic II item and the original NVPS were quite low. This result is congruent with the finding by Odhner et al that this item had the lowest correlation to the total scale.

Both the original and the revised NVPS showed significant differences between the ratings from before, during, and after the painful stimulus (original, 135.86, \(P<.001\), \(n=121\);

![Figure 1](https://example.com/figure1.png) Distribution of pain ratings.
revised, 145.05, $P < .001$, $n = 121$).

Figure 2 shows the mean ranks from the nonparametric Friedman’s repeated-measures test for each time for both scales. Post hoc testing showed that for both scales and every individual item, the pain rating during the intervention was significantly higher than the pain rating before or after the intervention with a single exception. Ratings on the physiologic II item did not increase significantly as expected during the painful procedure. Figure 3 shows the mean rank for this item compared with the respiratory item, which is the item that replaced it in the revised NVPS.

**Discussion**

Both total scales were supported for overall construct validity (Figure 2). However, as seen in Figure 3, the construct validity of the respiratory category is supported by the changes shown from before, during, and after, whereas the physiologic II category showed little variation. This result indicates that the physiologic II item was not discriminatory at an acceptable level. This finding supports the decision of Wegman’ to substitute the respiratory item in the revised version. Reliability of both versions was acceptable for the scores obtained during and after painful interventions. Internal consistency as indicated by the Cronbach $\alpha$ was a little better for the revised version. Although the original version had 1 item-to-total correlation that was slightly higher than in the revised version, the low item-to-total correlation of the physiologic II item also supported the superiority of the revised version. These findings tend to support the use of the revised NVPS rather than the original NVPS.

**Limitations**

The development of nonverbal pain instruments for sedated ICU patients receiving mechanical ventilation is at an early stage, and the complex clinical environment and patient conditions presented difficulties in measurement and in control.

**Table 3** Item to total correlations (Spearman correlation)

<table>
<thead>
<tr>
<th>Item</th>
<th>Face</th>
<th>Activity</th>
<th>Guarding</th>
<th>Physiologic I</th>
<th>Respiratory</th>
<th>Physiologic II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before intervention</td>
<td>0.835</td>
<td>0.554</td>
<td>0.303</td>
<td>0.291</td>
<td>—</td>
<td>0.219*</td>
</tr>
<tr>
<td>During intervention</td>
<td>0.708</td>
<td>0.657</td>
<td>0.639</td>
<td>0.557</td>
<td>—</td>
<td>0.257</td>
</tr>
<tr>
<td>After intervention</td>
<td>0.607</td>
<td>0.440</td>
<td>0.430*</td>
<td>0.724</td>
<td>—</td>
<td>0.277</td>
</tr>
<tr>
<td><strong>Revised</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before intervention</td>
<td>0.809</td>
<td>0.542</td>
<td>0.300</td>
<td>0.306</td>
<td>0.332</td>
<td>—</td>
</tr>
<tr>
<td>During intervention</td>
<td>0.648</td>
<td>0.588</td>
<td>0.663</td>
<td>0.577</td>
<td>0.689</td>
<td>—</td>
</tr>
<tr>
<td>After intervention</td>
<td>0.585</td>
<td>0.415</td>
<td>0.422</td>
<td>0.703</td>
<td>0.554</td>
<td>—</td>
</tr>
</tbody>
</table>

*Correlation is significant at the .05 level (2-tailed). All other correlations are significant at the .01 level (2-tailed). Dashes indicate not applicable.
in our study. The NVPS has previously been tested in only 1 other hospital. An additional area for concern is that the data collectors were aware of the stage (before, during, or after the intervention) when they completed their ratings. This knowledge could have influenced their scoring on the 3 observational items (face, activity, guarding) because they may have expected the scores to be higher during the intervention. This possible bias is potentially problematic because the main finding that supports the construct validity of the scale is that the scores increased significantly during the painful procedure. Psychometric consultation should be considered for studies that test complex, observational measurements.

**Implications for Practice**

Accurate assessment and pain management are essential to quality patient care. The Joint Commission standards of care underscore the importance of these practices. Our findings support the revised NVPS as a potentially valid and reliable observational tool for assessing pain in this ICU population of patients who are sedated and receiving mechanical ventilation but not paralyzed. Because the NVPS has been validated in only 1 other published study, the tool should be further tested in new ICU populations for validity and reliability.

Implementation of a new practice in any setting requires careful planning, staff involvement and motivation, training, and resources. For instance, it would be important to have laminated cards (Table 4) with the revised NVPS, or any nonverbal pain scale, at the bedside or taped to the head of the bed to remind nurses of the rating scale and the policy for pain assessment. Regular monitoring and feedback by a designated change agent and problem solver would be useful to maintain and institutionalize the practice until it becomes part of standard care.

**Implications for Future Research**

Additional research in ICU patients may also be useful to compare mean ratings of alert, observed patients on the NVPS with the corresponding self-reports of these same patients on a 0 (no pain) to 10 (most severe pain) numeric scale. Such research may be useful to test the construct validity of the tool further and might permit estimations of mild, moderate, or severe pain. It would also be useful to compare the NVPS with any other published nonverbal scales designed for use in ICU patients receiving mechanical ventilation.

Another promising observational tool for assessing pain in sedated ICU patients receiving mechanical ventilation, the BPS, is valid and reliable. Nurses using the BPS rate facial expressions from 1 (relaxed) to 4 (grimacing), upper limb movement from 1 (no movement) to 4

### Table 4 Revised card for Nonverbal Pain Scale

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1. Face</td>
<td>No particular expression or smile</td>
<td>Occasional grimace, tearing, frown or wrinkled forehead</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frequent grimace, tearing, frown or wrinkled forehead</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2. Activity (movement)</td>
<td>Lying quietly, normal position</td>
<td>Seeking attention through movement of slow cautious movements</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Restless activity and/or withdrawal reflexes</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3. Guarding</td>
<td>Lying quietly, no positioning of hands over areas of body</td>
<td>Splinting areas of the body, tense</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rigid, stiff</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4. Physiologic I (vital signs)</td>
<td>Stable vital signs, no change in past 4 hours</td>
<td>Change over past 4 hours in any of the following: SBP &gt;20, HR &gt;20, RR &gt;10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change over past 4 hours in any of the following: SBP &gt;30, HR &gt;25, RR &gt;20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RR &gt;10 above baseline or 5% ↓ SpO2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mild asynchrony with ventilator</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RR &gt;20 above baseline or 10% ↓ SpO2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Severe asynchrony with ventilator</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5. Respiratory</td>
<td>Baseline RR/SpO2</td>
<td>RR &gt;10 above baseline or 5% ↓ SpO2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complaint with ventilator</td>
<td>Mild asynchrony with ventilator</td>
<td>RR &gt;20 above baseline or 10% ↓ SpO2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Severe asynchrony with ventilator</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Revised Nonverbal Pain Scale (NVPS) Total</th>
<th>Score</th>
</tr>
</thead>
</table>

Abbreviations: HR, heart rate; RR, respiratory rate; SBP, systolic blood pressure; SpO2, oxygen saturation as measured by pulse oximetry.

*Reprinted with permission of Strong Memorial Hospital, University of Rochester Medical Center, Rochester, New York, developer and copyright holder of the scale.*
(fully retracted), and ventilator compliance from 1 (tolerant of movement) to 4 (unable to control ventilation), yielding ratings from 3 to 12. Our research council was uncomfortable with the absence of any physiological indicator on the BPS and also with the rating scale from 3 to 12 because nurses were accustomed to pain ratings from 1 to 10. The council members also thought that the use of 2 different scales, with different values, in the same unit would be confusing, even though they realized that observational ratings were not equivalent to self-reported intensity ratings. More recently, 2 additional validation studies13,14 of the BPS, in Moroccan and Australian populations, have been reported. The Moroccan study13 was not available, but the investigators in the Australian study14 assessed the validity of the BPS by comparing painful (repositioning) and not painful (eye care) procedures in ICU patients. The validity of the ratings was supported by descriptive findings that 73% of the BPS scores increased significantly during and after the painful procedures ($P < .003$), but only 14% of the scores increased after the eye care procedures ($P = .36$). Nurses selecting a pain scale for use in the ICU for sedated patients receiving mechanical ventilation might want to compare the NVPS and the BPS directly by testing these 2 scales in their patients. CCN

eLetters

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

Acknowledgments

We thank Jeff Allen, James Pullen, Pat Le, Carla Hohn, Jill Danahay, Wendy Siegel, Lori Conroy, and Michelle Rude for all their help collecting data. We also thank Kim Good for her efforts in data entry.

Financial Disclosures

None reported.

References


Further Validation of the Nonverbal Pain Scale in Intensive Care Patients
Anne Marie Kabes, Janet K. Graves and Joan Norris

Crit Care Nurse 2009;29 59-66 10.4037/ccn2009992
©2009 American Association of Critical-Care Nurses
Published online http://ccn.aacnjournals.org/

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

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

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

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

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