**Tool:** Pain Assessment for the Dementing Elderly (PADE)  
**Tool developers:** Villaneuva, M.R., Smith, T.L., Erickson, J.S., Lee, A.C., Singer, C.M.  
**Country of origin:** USA

### Conceptualization

**Panel rating:** 1

<table>
<thead>
<tr>
<th>Purpose</th>
<th>The Pain Assessment for the Dementing Elderly (PADE) is a tool for assessment of pain in individuals with advanced dementia. It was developed to help caregivers assess patient behavior that may indicate pain.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual basis</td>
<td>The terms distress and pain are used interchangeably. Association between pain and agitation was noted.</td>
</tr>
</tbody>
</table>

Pain assessment by observation rests upon three key assumptions:
1. Facial characteristics, body posture and movement patterns can indicate the presence of pain.
2. Pain can interfere with activities of daily living (ADLs) such as dressing and eating.
3. Caregivers can reliably observe and rate behavior.

Assumptions reflect a conceptual basis for pain by observation of behaviors.

### Item Generation

**Tool items**

- 24 items in three parts:
  - I Physical
    - observable facial expression
    - breathing pattern
    - posture
  - II Global Assessment – proxy evaluation of pain intensity
  - III Functional (ADL)
    - dressing
    - feeding oneself
    - transfers from wheelchair to bed

**Item generation process**

Item generation is based on:
- literature review (Kovach: facial features and ADL)
- interviews with nursing staff at care facilities
- naturalistic observation by research associate (logged 30 hours over a 2 week period in a dementia care unit).

Items were validated by clinical experts in three settings. Feedback was provided by staff regarding ease of use, comprehensiveness and manual instructions. Adjustments were made.

### Content Validity

The PADE has not been subjected to content validation by independent experts in pain in elders with dementia.

### Panel Commentary

Conceptually the tool includes 5 categories of pain indicators noted in the AGS Persistent Pain Guidelines. However, operationalization of specific indicators is not clearly supported.

The assumption that caregivers can reliably rate the intensity of elders’ pain is not supported by current literature.

Most items in the tool measure intensity of behaviors, not severity of pain, which may be appropriate; however, the relationship between behavior
intensity and pain has not been established in this population.

Construction of the scale is problematic in its development in several ways, based on available information:
1. Three parts to the tool are described, yet 24 individual components are presented without clarity as to which domain of pain they are measuring.
2. Relevance of some indicators to assessment of pain is lacking (e.g. Neatness of grooming).
3. Caregiver judgment of pain severity as the gold standard has not been substantiated.
4. There is inconsistency in narrative description of the tool and the tool illustration in the appendix, e.g. The narrative documents Likert rating 1-4, but the tool illustrates a semi-VAS format.
5. Items 15-24 related to functional ADL are assessed retrospectively from the resident’s chart. However, all other items are rated based on the resident’s current situation. Differences in timing of assessment components could be problematic.
6. The relationship of the ADL section in documenting pain is unclear. However, if the tool is used regularly and consistently, it may potentially show changes over time (e.g. percent of time out of bed, percent of time awake, amount of meals eaten) that could be potential indicators of pain impact.

Although the tool was presented for validation by clinical experts, little information is available as to their education and clinical background. The results regarding tool comprehensiveness and ease of use are not presented. There is also no information as to whether these clinical experts were independent from those staff members involved in the development of the initial set of pain items.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Panel rating: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1: Conducted to establish reliability</td>
<td></td>
</tr>
<tr>
<td>Setting: 4 long term care facilities including 3 skilled nursing facilities and a locked dementia assisted-living facility.</td>
<td></td>
</tr>
<tr>
<td>Subjects: 25 LTC residents with advanced dementia</td>
<td></td>
</tr>
<tr>
<td>Average age: 84 years</td>
<td></td>
</tr>
<tr>
<td>Gender: Male: 9, Females: 16</td>
<td></td>
</tr>
<tr>
<td>Dementia Assessment: Global Deterioration Scale: Average score: 5.60 (±0.71), Range 4-7.</td>
<td></td>
</tr>
</tbody>
</table>

Pain conditions: 60% suffered potentially painful medical conditions and 80% had prescribed analgesics, 55% prn and 45% scheduled dose.

Residents were selected by the RN manager based on which caregivers were involved in the study. Most raters were certified nursing assistants or similarly trained caregivers.

Study 2: Conducted to replicate reliability finding of study 1 and to assess validity |
| Setting: One long term care facility in Washington state |
| Subjects: 40 residents with similar level of dementia |
| Average age: 81.33 years (±7.67), Range 66-92 years |
| Gender: Male: 8, Female: 32. |

Dementia assessment: Global Deterioration Scale: Average score: 5.28
Pain conditions: 25% suffered potentially painful medical conditions and 27.5% had prescribed analgesics, 9% prn only and 91% scheduled dose.

**Panel Commentary**

Focus on long term care setting is clearly identified. The subjects are clearly identified as having severe dementia using appropriate measurements tools.

Residents were selected by the RN manager based on which caregivers were involved in the study, which may result in selection bias. Little information is available on caregivers who participated and how they were chosen.

Age is appropriate.
There is gender balance.
No information regarding racial/ethnic diversity is available.

Using 5 subjects per item as a rule of thumb, a sample size of 120 subjects (24 items x 5 subjects per item) would be needed. Thus, with 25 and 40 subjects in study 1 and 2 respectively the sample sizes are insufficient.

### Administration, Scoring, Feasibility

**Panel rating: 1**

- **Administration, Scoring, Feasibility**

  The tool has 24 items in three parts (see “Item Generation” section above).

  Items 1-12, 14 and 22-24 are rated using a Likert scale. This is scored 1-4 (based upon equally spaced division on the line). Items 13 and 15-21 are multiple choice, scoring 1-4.

  The method of administration is described on the tool and raters are referred to a tool manual. The instructions on the tool separate the items into two major sections. However, these do not correspond to the above mentioned three part division.

  For items 1-14 the rater is instructed to observe the resident’s behavior for five minutes and then score the items by placing a vertical slash mark on a visual analogue scale. For items 8-10 the rater is instructed to mark the item “Not Applicable” if the resident is silent during the observation period. Item 13 is multiple choice.

  The instructions to questions 15-24 regarding functional ADL indicate that the questions should be answered by reviewing chart documentation. Instructions ask the rater to review the Master Sheet in the manual for directions regarding how to answer the questions. Most questions are based on chart documentation from the previous 24 hours.

  There is no indication in the research report or on the tool form as to how to proceed once the individual items are scored. There is no indication of a score related to each of the three main parts of the tool or total score for the assessment.

  There is no guide as to how to interpret the results of the assessment.

  Ratings were conducted by CNA or personnel with similar training. One hour of training was provided to raters. The PADE required 5-10 minutes to complete. An RN supervised the first two ratings to check for adherence to protocol.

**Panel Commentary**

Several issues are identified related to tool administration and scoring:
1. Three parts to the tool are described, yet 24 individual components are presented without separation into parts and without clarity as to which part of the tool the individual items belong.
2. There is an expectation of finding data for some items in the patient chart. However, based on current documentation practice, accurate information may not be available.
3. Scoring procedures are very confusing due to inconsistency in narrative description of the tool and the tool representation in the appendix, eg. The narrative documents likert rating 1-4, but the tool illustrates a semi-VAS.
4. A variety of scaling approaches has been used (eg. VAS items; Likert 1-4 scales), however, all with different anchors without clear reason for inconsistent scaling approaches. This is confusing and contributes to complexity of interpretation.
5. A Score of 0 is given if an item is marked as “Not Applicable”. The impact of this on the overall result and underestimation of pain is not addressed.
6. Caregiver judgment of pain severity as the gold standard is not substantiated in the literature.
7. Interpretation of tool score is unclear.

An instruction manual has been developed for the tool. Further explanation or clarification regarding administration and scoring of the tool may be documented. However, the instruction manual was not available for this review.

**Clinical utility**
- **Time**: It is suggested that the tool administration takes 5-10 minutes. However, data to support this is not provided. The rater is asked to observe the subject for five minutes before rating. Moreover, considering the complexity of the items, the variety of scaling approaches and the expectation of finding answers in the patient record, this tool may take considerably longer to administer than the suggested 5-10 minutes.
- **Skill needed**: The raters in the research report are primarily nursing assistants. However, there is no discussion related to nursing assistants’ scope of practice.

### Reliability

<table>
<thead>
<tr>
<th>Panel rating</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Study 1</strong></td>
<td>Each resident (N=25) (see sample characteristics for study 1 above) was assessed by 2 independent observers over 10 days (not consecutive). A total of 784 observations were made.</td>
</tr>
<tr>
<td></td>
<td>Ratings were made by caregivers, most of whom were certified nursing assistants or similarly trained caregivers.</td>
</tr>
<tr>
<td></td>
<td>The caregivers received a 1-hour training session.</td>
</tr>
<tr>
<td></td>
<td>A registered nurse was chosen for onsite management and contact with the investigators. The RN observed the first two ratings to check for adherence to protocol.</td>
</tr>
<tr>
<td></td>
<td>Part I: alpha 0.77</td>
</tr>
<tr>
<td></td>
<td>Part III: alpha 0.63</td>
</tr>
<tr>
<td><strong>Study 2</strong></td>
<td>Each resident (N=40) (see sample characteristics for study 2 above) was</td>
</tr>
</tbody>
</table>
simultaneously assessed by 2 independent raters at time 1 and time 2. A total of 160 observations were made.
Ratings were made by caregivers, most of whom were certified nursing assistants or similarly trained caregivers.
The caregivers received a 1-hour training session.
A registered nurse was chosen for onsite management and contact with the investigators. The RN observed the first two ratings to check for adherence to protocol.

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I: alpha</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Part III: alpha</td>
<td>0.24</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**Interrater reliability**

**Study 1**
(See description of study 1 under Internal Consistency above.)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I ICC</td>
<td>0.93</td>
<td>(.92, .94)</td>
</tr>
<tr>
<td>Part II ICC</td>
<td>0.81</td>
<td>(.77, .85)</td>
</tr>
<tr>
<td>Part III ICC</td>
<td>0.96</td>
<td>(.95, .96)</td>
</tr>
</tbody>
</table>

**Study 2**
(See description of study 2 under Internal Consistency above.)

<table>
<thead>
<tr>
<th></th>
<th>Estimate (95% CI)</th>
<th>Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I ICC</td>
<td>.93 (.87, .96)</td>
<td>.95 (.91, .97)</td>
</tr>
<tr>
<td>Part II ICC</td>
<td>.89 (.80, .94)</td>
<td>.54 (.13, .77)</td>
</tr>
<tr>
<td>Part III ICC</td>
<td>.94 (.83, .94)</td>
<td>.93 (.88, .96)</td>
</tr>
</tbody>
</table>

**Test-retest reliability**

**Study 1**
(See description of study 1 under Internal Consistency above.)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I: ICC</td>
<td>.70</td>
<td>(.47-.83)</td>
</tr>
<tr>
<td>Part II: ICC</td>
<td>.34</td>
<td>(.16-.63)</td>
</tr>
<tr>
<td>Part III: ICC</td>
<td>.89</td>
<td>(.81-.94)</td>
</tr>
</tbody>
</table>

**Study 2**
(See description of study 2 under Internal Consistency above.) The time interval for test-retest was 20 days.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I: ICC</td>
<td>.98</td>
<td>(.97-.99)</td>
</tr>
<tr>
<td>Part II: ICC</td>
<td>.70</td>
<td>(.54-.81)</td>
</tr>
<tr>
<td>Part III: ICC</td>
<td>.98</td>
<td>(.97-.99)</td>
</tr>
</tbody>
</table>

**-Panel commentary**
The characteristics of the raters collecting data is not clear (age, clinical background, level of education). Supervision/monitoring of data collection by caregivers appears limited which may explain some of the variability in test results.

**General study I**
Although a total of 784 observations over 10 days for 25 residents was made in study 1, the number of observations included in each of the reported results is unclear. 784/2 raters/25 residents means that each resident was observed on average 15 times, eg. more than once per day. What situation triggered the observation? With several observations per day, are the observations really independent, especially since function items are based on last 24 hours of chart documentation?
Internal consistency
It is unclear what data were entered into analysis for items with Likert scales. In study 1, alpha measures were acceptable. In study 2, alpha measures for part I were acceptable. However, for part III the alphas were poor at both time points.

Preliminary internal consistency of the tool is not well established for all components.

Interrater reliability
Raters were nursing assistants.
Intraclass correlations are appropriate test for data.
Reliabilities are mostly strong across study 1 and 2.

Test-retest reliability
In study 1 the time interval for the two measures included in the test-retest assessment is not reported, thus appropriateness of the procedure cannot be determined.

In study 2 the interval was 20 days (but this interval varied slightly). No control over change in pain condition resulting from analgesic or other intervention is noted.

The intraclass correlations for part I and III are within acceptable range in study 1 although the correlations for part I are in the lower acceptable range and the 95% confidence interval is wide suggesting variability. Intraclass correlations for part II are low.

The intraclass correlations are acceptable for study 2 for all three parts. However, correlations for part II are in the lower acceptable range with wide confidence intervals.

Validity: Criterion or construct
Panel rating: 1

Construct validity/ Criterion related validity
A description of subjects for each of the two studies is provided under Subjects above. A description of study procedures is provided under Internal consistency above.

Construct validity
The PADE was compared with Cohen-Mansfield Agitation Inventory (CMAI). The CMAI has three subscales which assess different subtypes of agitation: verbal, physically aggressive and physically nonaggressive.
It was hypothesized that the PADE would correlate with verbal but not physically aggressive agitation.

Convergent validity
The PADE was administered at T2. The CMAI was administered at T2 and refers to the patient’s behavior over the previous 3 weeks.

Results:

<table>
<thead>
<tr>
<th></th>
<th>CMAI Aggressive</th>
<th>CMAI Nonaggressive</th>
<th>CMAI Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>PADE Part I</td>
<td>r=.21</td>
<td>r=.08</td>
<td>r=.30</td>
</tr>
<tr>
<td>PADE Part II</td>
<td>r=.13</td>
<td>r=.18</td>
<td>r=.14</td>
</tr>
<tr>
<td>PADE Part III</td>
<td>r=.40</td>
<td>r=.40</td>
<td>r=.42</td>
</tr>
</tbody>
</table>
Criterion validity

Study 2: (N=40) (see sample characteristics above).
Criterion validity was assessed by comparing residents (1) with or without painful conditions (2) for whom pain was or was not a significant clinical factor, and (3) residents with or without psychoactive medications.

1) Residents with or without painful conditions
Using chart review to classify the 40 residents, 10 were judged to suffer and 30 not to suffer from painful conditions by an independent RN. There were no statistically significant differences between groups with or without painful conditions on the PADE subtests or the CMAI.

2) Residents for whom pain was or was not a significant clinical factor
The sample was divided into two groups based on whether the independent RN judged pain to be an ongoing, significant clinical factor in resident’s care: (n=8 pain significant factor and n=32 pain not a significant factor).
Three CMAI subscales and the PADE Parts I, II and III were used to compare the groups. The pain as a significant clinical factor group had statistically significantly higher scores on the following subscales:

- CMAI Verbal  \( z=-2.83 \) (p=0.01)
- PADE Part I  \( z=-2.31 \) (p=0.01)
- PADE Part II  \( z=-4.30 \) (p<0.001)
- PADE Part III  \( z=-2.31 \), p<0.01)

3) Residents with or without psychoactive medications
Residents were divided into 2 groups based on whether there were presently taking psychoactive medication or not. The group on prescribed psychoactive medications had statistically significantly higher scores on the following subscales:

- CMAI Verbal  \( z=-1.95 \) (p=0.05)
- CMAI Aggressive  \( z=-3.34 \) (p=0.01)
- PADE Part I  \( z=-3.19 \) (p<0.01)
- PADE Part III  \( z=-3.25 \) (p<0.01)

-Panel commentary
The classification of subjects into groups rests on the knowledge of one independent rater.
Validity testing has not established the usefulness of the PADE in identifying those with and without pain conditions, suggesting relevant pain related behaviors may not be present. Preliminary data may suggest those with significant pain problems may be identified through use of the tool. Although agitation and pain have been associated in elders with dementia, validation of this relationship should complement examination of other validity constructs.
Usefulness of the PADE Part II is not supported based on psychometric evaluation conducted to date.
Although use of psychoactive medication differentiates groups, control of agitated behavior by medication could impact the behavioral presentations of pain and agitation/aggression.
Use of analgesics and psychoactive medication were not controlled in relation to timing of assessments and may impact results.

Summary of panel evaluation of pain assessment tool
The PADE was developed to provide a simple tool for assessing pain in individuals with advanced dementia. Although 5 categories of indicators from the AGS guideline are included, operationalization of the tool shows inconsistencies in construction, presentation and lack of clarity in scoring and interpretation. The relevance of some indicators to assessment of pain is lacking. The tool assumes caregivers can judge pain severity from behavioral observation which has not been demonstrated in this population.
Lack of control of factors that could impact the tool performance (eg. receipt of analgesics and psychotropics) limits interpretation of data.

Because assessment activities are outside the scope of nursing assistant practice, it will be important to determine if the expectations of the tool for NA’s are actually screening activity.

Although a relationship between pain and agitation has been demonstrated, the CMAI as the only gold standard for comparison is insufficient for judgment of tool validity. The PADE requires clarification and further revision/testing.

Source of evidence

Key to panel rating
3= Available evidence is strong
2= Available evidence supports need for further testing
1= Available evidence is insufficient and/or tool revisions are needed
0= Evidence is absent

Evaluation completed by:
K. Herr, S. Decker, K. Bjoro, University of Iowa.
Contact information: keela-herr@uiowa.edu