MS MARGARET ANN FOX (Orcid ID : 0000-0002-7971-6364)

Article type   : Original Article

[Original article: 4 tables]

Self-report of pain in young people and adults with spastic cerebral palsy: interrater reliability of the revised Face, Legs, Activity, Cry, and Consolability (r-FLACC) scale ratings

MARGARET A FOX[1,2]
RITA AYYANGAR[3]
REBECCA PARTEN[1]
HEIDI J HAAPALA[1]
STEPHEN G SCHILLING[4]
CLAIRE Z KALPAKJIAN[1]

1 Michigan Medicine, Department of Physical Medicine and Rehabilitation, Ann Arbor, MI; 2 School of Nursing, University of Michigan, Ann Arbor, MI; 3 Michigan Medicine, University of Michigan, Department of Physical Medicine and Rehabilitation and, Pediatric Palliative Care Medicine, Ann Arbor, MI; 4 Survey Research Center, Institute for Social Research, Ann Arbor, MI, USA.

Correspondence to Margaret A. Fox, School of Nursing, University of Michigan, 325 E. Eisenhower Parkway, Ann Arbor, MI 48108, USA. E-mail: mafox@med.umich.edu

PUBLICATION DATA

ABBREVIATIONS

FLACC   Face, Legs, Activity, Cry, and Consolability

r-FLACC   Revised Face, Legs, Activity, Cry, and Consolability

NRS   Numeric rating scale

AIM People with cerebral palsy (CP) are often unable to express pain owing to cognitive or speech impairments. Reports that rely on observation can be inaccurate, because behaviours such as grimacing, common in people with spastic CP, resemble pain expressions. We examined preliminary validity and reliability of the revised Face, Legs, Activity, Cry, and Consolability (r-FLACC) scale in persons with spastic CP.

METHOD Forty-eight young people and adults (35 females, 13 males; mean [SD] age 29y 2mo [13y]) were video-recorded during a standard examination, rating their pain (0–10) afterwards. Two raters completed the r-FLACC using the video recordings. Interrater reliability was assessed with an unconditional cross-classified random effects model and Item Response Theory approach; Pearson correlations measured agreement between raters and participants.

RESULTS Mean (SD) participant (n=48) pain scores were 2.48 (2.5) and mean (SD) r-FLACC scores were 1.46 (1.68). There was moderate agreement between raters (intraclass coefficient 0.41 and 0.57 respectively) but low agreement between participants and raters (r=0.26). There were no significant effects for raters (lay observers, nurses, physicians, and inexperienced raters).

INTERPRETATION Results provide mixed support for the interrater reliability of the r-FLACC in people with spastic CP.

What this paper adds

- The revised Face, Legs, Activity, Cry, and Consolability (r-FLACC) scale can be reliably used by experts and lay raters for persons with spastic cerebral palsy (CP).
- Support is mixed for interrater reliability of the r-FLACC scale used with persons with spastic CP.

[Main text]

Pain is a common problem in cerebral palsy (CP), with approximately 62% of children and 75% of adults with CP reporting chronic pain, far more than their peers who are without disabilities.[1,2] In one study, 85% of 442 young adults reported pain, with the majority reporting greater than moderate pain.[3] People with CP report pain during daily tasks such as dressing, transfers, and range of motion, and may undergo painful medical interventions.[4] Pain-assessment scales are predominantly based on self-report and use a numeric or visual analogue scale because pain is a participative experience and self-report is the criterion standard.[5,6] Alternative self-report measures such as pictures or the use of adaptive communication devices are preferred whenever possible for young people and adults with communication or cognitive impairment who cannot use standard self-report measures.[2,5,7–9] Pain report by proxy is often used; however, given the subjective nature of pain, a proxy may not have accurate insight into an internal experience of pain and may over- or underestimate pain.[10] In a large study of 1174 young people with CP, a parent proxy rating of pain had a moderate association with self-report, but pain was overestimated if it was infrequent and mild and underestimated if frequent or severe.[4] In a study of 252 children and young people with CP, 47% of children and young people reported some pain versus 55.6% of their caregivers and 38.7% of physicians.[11] To assess pain in those who are unable to verbally communicate pain severity, observational tools have been developed to interpret pain behaviour.[6,12,13] The use of observational tools, however, can be a challenge in a

population such as those with CP where some common observations such as flexion and grimacing can be interpreted as pain behaviours but are actually characteristics of spasticity. Other barriers to utilizing observational assessment tools include time to train the observer and to administer the tool in a clinical encounter. For example, the Non-Communicating Children Pain Checklist and Non-Communicating Adult Pain Checklist have been developed and extensively evaluated for its sensitivity in clinical settings.[6,13,14] However, specific training of the rater is required and a minimum 10-minute observation period is also necessary, which may be too lengthy for a screening tool in a time-pressed medical setting when pain is only one of many pressing issues.[14] The 20-item, relatively quicker at 5 minutes, observation Pediatric Pain Profile has been validated and found to be reliable for use in children with cognitive impairment. The Pediatric Pain Profile is used by caregivers and health professionals to establish baseline pain scores on a 'good day' and rescore when the child is experiencing pain. Some of the 20 items measure motoric behaviours that may be present in the person with spastic CP on a good day, such as grimacing, teeth grinding, tenseness, flexion, involuntary movements, or spasms.[15] Other observational and proxy measures will be useful for further evaluation if the presence of pain is suspected.[16]

The Face, Legs, Activity, Cry, and Consolability (FLACC) scale was developed to observe signs of pain in preverbal children, but it is also used with non-verbal people. One limitation of the use of the FLACC in persons with CP is that behaviours in the scale, such as 'legs drawn up' and 'arched' are common positions observed in persons with spastic CP, even without pain being present. To address this concern and improve the reliability and validity of the tool in people with cognitive impairment and communication problems, the revised FLACC (r-FLACC) was developed.[17] The r-FLACC scale adds descriptions of pain provided by parents and caregivers of children, which help distinguish pain from spasticity in someone who has chronically spastic muscles. The r-FLACC and Individualized Rating Scale can be customized indicating behaviours specific to each person, an important feature in people with spastic CP and cognitive impairment. Moreover, the r-FLACC can be completed in a very short time and is easy to use in busy healthcare settings by providers who are not experts in pain assessment.[18,19] To address a gap in tools for persons with CP who are unable to communicate their pain, we examined the interrater reliability of the r-FLACC in a sample of young people and adults with CP who were able to communicate their pain ratings.

METHOD

Participants

The study was conducted in an outpatient hospital clinic in a tertiary medical centre where children and adults are seen for disabling conditions, including CP, between September 2014 and June 2015. Participants were recruited at their regularly scheduled visit after a brief screen of their medical records to determine eligibility. Inclusion criteria were age 13 to 66 years with a diagnosis of spastic CP by a physiatrist; and the ability to communicate verbally, using signs, or with the use of a communication device. Participants were screened using the Communication Function Classification System to ensure they were able to communicate effectively with strangers (Communication Function Classification System levels I–III were eligible).[20] After the visit, a $10 gift card was mailed to the participant. A sample size of 48 was determined based on feasibility in terms of clinical service flow and study resources. Additionally, a sample of 48 was equally divisional by the six raters and was felt to be sufficient for relevant statistical analysis.

Four raters were healthcare specialists (two registered nurses and two physicians) and two were laypersons with personal or professional experience related to children with disabilities. All of the raters had at least some previous exposure to the FLACC and r-FLACC, and some had used the scale in clinical practice for many years. Standardized training on the use of the FLACC and r-FLACC was provided to each rater irrespective of experience or expertise. This study was approved by the Institutional Review Board of the University of Michigan Medical School (HUM00085574) and all participants and raters signed a written informed consent document.

Pain measurement

The r-FLACC is an observational pain measure based on the extensively evaluated and widely used preverbal pain assessment FLACC tool.[16,17,21] In 2002, Voepel-Lewis et al. tested the reliability and validity of the FLACC scale in children with cognitive impairment and described limitations in the use of the tool.[22] Revisions made to each category were based on parent and caregiver interview of pain behaviours characteristic of children with cognitive impairment. For example, in the face category the descriptor *appears sad or worried* was added to the mild pain section and *distressed looking face; expression of fright or panic* was added to the moderate-to-

severe section. In addition, in the revised version, parents and caregivers are offered additional descriptions and an open-ended blank to fill in individualized moderate-to-severe pain behaviour, resulting in improved reliability and validity in children with cognitive impairment.[17] Intensity is assigned a value on a 3-point scale (0–3) for each of five categories: face, legs, activity, cry, and consolability. For example, the face category scores include 0=no particular expression or smile; 1=occasional grimace/frown, withdrawn or disinterested; and 2=consistent grimace or frown, frequent/constant quivering chin, clenched jaw. Total scores range from 0 to 10, with 10 indicating the worst pain.

Data collection and rater assignment

Firstly, participants were familiarized with the numeric rating scale (NRS) pain score from 0 (indicating no pain) to 10 (indicating worst pain) and they were informed that they would be asked immediately after the exam to report any pain experienced during the exam. Secondly, the participants were oriented to the r-FLACC form with the five categories and descriptions of pain behaviours. The participant, research staff, and/or the participant's family/staff facilitated this task by providing assistance as needed, such as reading the behaviours aloud and recording individualized behaviours described by the participants. Once the unique pain descriptions were documented, a standard examination including tone and range of motion was video recorded by a research assistant. The standard clinician examination includes evaluation of muscle tone and joint range of motion evaluation and was not meant to provoke pain, but often people with spasticity do express at least some discomfort during the examination. Hadden et al. report measuring pain in children during routine stretch procedures in physiotherapy sessions.[10] Immediately after the examination, and after the video recorder was stopped, the research staff asked the participant to report the pain level experienced during the preceding examination on a NRS of 0 (no pain) to 10 (worst pain).

Sound was recorded during the examination but removed during later editing so the rater would not have any incidental description of pain that may have taken place between clinician and participant. Videos were then assigned to raters (16 participants for each rater, two raters for each participant) using block randomization. In the videos, the emphasis was on participant facial expressions and other non-verbal behaviours, but the view was sometimes obscured by the position of the participant and recorder in the small examination room. The target length of the

video was about 1 minute. Actual video lengths ranged from 51 seconds to 2 minutes and 44 seconds. Video recordings were transferred from the camera to secure servers at the end of each day. Each video recording was edited before review by raters (e.g. audio removed, unnecessary footage removed, such as pre-examination preparation) using Adobe Premiere Elements 13 (Adobe Systems, Mountain View, CA, USA). If a recording appeared too long, the principal investigator (MAF) determined content that could be removed to reduce the overall length to ensure no content was removed that would have provided the raters with important information.

Statistical analysis

We were interested in overall differences in participant versus rater pain scores and agreement between raters or interrater reliability. Firstly, we used descriptive statistics to describe score differences between participants and raters. Secondly, we examined interrater reliability of raters, irrespective of participant scores, with two methods. The first method fit an unconditional cross-classified (i.e. raters were crossed with and not nested within participants) random-effects model using MLwiN 3.0 (Centre for Multilevel Modelling, University of Bristol, UK), with random effects for participant, rater, and observation (i.e. pain score by the rater).[23] For the first method, pain ratings were treated as continuous. Estimates of the variance of the random effects were used to calculate the intraclass coefficient. To examine whether the type of rater was associated with higher or lower pain scores, we then fit random effects models where rater effects were allowed to vary by type of rater. Specifically, we examined the effects for lay raters, nurses, and physicians, and then inexperienced raters versus experienced raters, where the experienced physicians and nurses were the experienced raters. The second method to examine rater agreement was an Item Response Theory approach where each rater score was considered an item, which allowed us to account for the ordinal nature of the pain ratings. A graded response model was then fit to the pain scores using IRTPRO 2.1 (Scientific Software International, Lincolnwood, IL, USA), with item discrimination fixed across the raters.[23] The reliability of the resulting Item Response Theory scale scores was then estimated.[24] This gave an estimated reliability for the average score using all raters – the reliability using a single rater could then be estimated using the Spearman–Brown prophecy formula. Lastly, we calculated the correlation between raters and participant pain ratings.[23]

RESULTS

Participant characteristics

In total, 277 patients with upcoming clinical appointments were screened in medical records; of these, 79 were eligible and 48 were enrolled. Of the 31 participants who were eligible but not enrolled, 11 did not keep their clinic appointment, 13 were not approached because the clinic was too busy, six were not interested, and one was enrolled but was not video recorded because the camera battery failed. The sample was mostly female, around 30 years old, with the majority having spastic quadriplegia, and half in Gross Motor Function Classification System Level IV. See Table I for sample characteristics. A sample of eligible non-participants was similar to the participants, with an average age of 31 years, and the majority diagnosed in Gross Motor Function Classification System Level IV with spastic quadriplegia.

Participant and rater pain ratings

All participants completed the r-FLACC and 28 of 48 added unique pain characteristics, predominantly in the face or activity categories (see the summary in Table II). There were 96 pain ratings (two per participant) from the six raters. The participant NRS scores averaged 2.48 (SD 2.5) and rater scores averaged 1.46 (SD 1.68). Slightly over half of r-FLACC ratings were higher than self-report and a third were lower. See Table III.

Interrater reliability

For the unconditional random-effects model, variance of the estimated random effects for participant, rater, and observation were 1.25, 0.24, and 1.55 respectively, yielding an intra-class coefficient of 0.41 (moderate agreement). For the Item Response Theory model, the estimated interrater reliability was slightly higher at 0.57 (moderate agreement), likely owing, in part, to the rater scores being treated as ordinal. Different types of raters did not differ with respect to their average rating, as Table IV shows that there were no statistically significant effects for lay raters, nurses, physicians, and inexperienced raters. Finally, agreement between rater scores and participant scores was 0.26 (small).

Rater open-ended comments

There were 39 comments made by the raters on their respective r-FLACC forms. These were grouped into three themes: explanation of their rating, view of the participant, and lack of sound on the recording. Twenty-four comments explained or justified the rating they gave, described the observation, why the rating they chose could have been different or commented on their own assessment as they watched the videotape. Examples include 'Several facial movements – not clear if dystonia or brief grimace' or 'I noted the participant's eye had rapid movement when the examiner pushed on the hands to release spasticity'. Ten comments cited the limited view of the participant in the video that was sometimes obscured by the patient's position as exemplified in this comment: 'could not see participants face for most of the assessment' and 'difficult to assess face given location and partially covered by patient's own arm'. Five comments by raters indicated the lack of sound in the video was a hindrance in accurately rating the participant's pain. Study design was purposefully deleted and did not review the sound recorded in the exam room as the participants were in an evaluation with their healthcare provider. We did not ask the participant or the examiner to discuss pain, omit discussion about pain or alter their conversation in any way. Including sound with words on the video recording would have negated our effort, but including sound without words would have been helpful to one rater, who commented, 'I think she whimpered even though I could not hear'.

DISCUSSION

The results of this study do not strongly support the use of the r-FLACC as a useful tool for assessing pain intensity in people with spastic CP and chronic pain. While interrater reliability was moderate, agreement between rater r-FLACC scores and participant ratings was modest. More than half of r-FLACC scores were lower than the participants' scores. While it can be argued that some of this discrepancy could be attributable to the different methods of arriving at a score (self-rating of 0–10 vs score on a series of items), the fact that in slightly more than half of ratings r-FLACC scores were lower than participant self-ratings suggests that the r-FLACC may be limited in approximating pain intensity. Moreover, categories that used pain behaviours such as flexion and grimacing, which are common signs in spastic CP, would tend to raise the r-FLACC score, not lower it, which is in contrast to our findings. One of the advantages of the r-FLACC is sensitivity to atypical pain characteristics owing to individualization of the scoring criteria by others who know the individual well, perhaps overcoming the base criteria, including

flexion, grimacing, and other signs common with high muscle tone. Other commonly used tools such as the Pediatric Pain Profile and the Non-Communicating Adult Pain Checklist include numerous descriptions of spasticity such as grimace, flexed inward, stiff spastic, and tense descriptors that cannot be distinguished from pain.[14,15]

Limitations

The r-FLACC was designed and validated for use in preverbal children with acute pain and was not created for young people and adults with chronic pain or spastic CP. Adults and children express pain differently and acute and chronic pain may be expressed differently. Also people who can communicate pain verbally, with a device or signs may have different characteristics of pain expression. Without further study, the validity of the r-FLACC in the spastic CP population may be limited. Ratings may have instead been negatively affected by the use of video versus in-person observation. Some information was lost using video recording rather than real-time observation and the view was sometimes obscured. It would have been useful to have a constant video of the face with a second camera. Both the r-FLACC and the NRS have a score ranging from 0 to 10, but they arrive at the score using very different methods. The r-FLACC score is limited to a maximum of 2 points in a given category and is not as fluid as the 0 to 10 NRS in rating a level of pain. The r-FLACC structure into categories encourages careful observation of the complete physical person and their behaviours, whereas the NRS is a rapid check of a person's self-report of pain intensity. While the r-FLACC is a strictly observational tool, and therefore unable to be transformed as a self-report scale, there may be other measures more suitable for investigations of its utility in this population.

CONCLUSION

The results suggest that the r-FLACC may not be a useful tool for accurately assessing pain intensity in people with spastic CP and chronic pain who are unable to verbalize their pain. The r-FLACC may have some applicability as a first step in an assessment to indicate the presence of pain and unique pain behaviours but may need to be followed by an in-depth pain assessment for intensity by a clinician. Future work to further development of pain-rating tools for non-verbal persons with CP is needed.

ACKNOWLEDGMENTS

REFERENCES

1. Van Der Slot WM, Nieuwenhuijsen C, Van Den Berg-Emons RJ, et al. Chronic pain, fatigue, and depressive symptoms in adults with spastic bilateral cerebral palsy. *Dev Med Child Neurol* 2012; **54**: 836–42.

2. Ramstad K, Jahnsen R, Skjeldal OH, Diseth TH. Characteristics of recurrent musculoskeletal pain in children with cerebral palsy aged 8 to 18 years. *Dev Med Child Neurol* 2011; **53**: 1013–8.

3. Sienko SE. An exploratory study investigating the multidimensional factors impacting the health and well-being of young adults with cerebral palsy. *Disabil Rehabil* 2018; **40**: 660–6.

4. Parkinson KN, Dickinson HO, Arnaud C, Lyons A, Colver A; SPARCLE Group. Pain in young people aged 13 to 17 years with cerebral palsy: cross-sectional, multicentre European study. *Arch Dis Childhood* 2013; **98**: 434–40.

5. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res* 2011; **63**(Suppl. 11): S240–52.

6. Lotan M, Ljunggren EA, Johnsen TB, Defrin R, Pick CG, Strand LI. A modified version of the non-communicating children pain checklist-revised, adapted to adults with

intellectual and developmental disabilities: sensitivity to pain and internal consistency. *J Pain* 2009; **10**: 398–407.

7. Vetter TR, Bridgewater CL, McGwin G Jr. An observational study of patient versus parental perceptions of health-related quality of life in children and adolescents with a chronic pain condition: who should the clinician believe? *Health Qual Life Outcomes* 2012; **10**: 85.

8. Walsh M, Morrison TG, McGuire BE. Chronic pain in adults with an intellectual disability: prevalence, impact, and health service use based on caregiver report. *Pain* 2011; **152**: 1951–7.

9. Voepel-Lewis T, Malviya S, Tait AR. Validity of parent ratings as proxy measures of pain in children with cognitive impairment. *Pain Manag Nurs* 2005; **6**: 168–74.

10. Hadden KL, LeFort S, O'Brien M, Coyte PC, Guerriere DN. A comparison of observers' and self-report pain ratings for children with cerebral palsy. *J Dev Behav Pediatr* 2015; **36**: 14–23.

11. Penner M, Xie WY, Binepal N, Switzer L, Fehlings D. Characteristics of pain in children and youth with cerebral palsy. *Pediatrics* 2013; **132**: e407–13.

12. Voepel-Lewis T, Zanotti J, Dammeyer JA, Merkel S. Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. *Am J Crit Care* 2010; **19**: 55–61.

13. Breau LM, Burkitt C. Assessing pain in children with intellectual disabilities. *Pain Res Manag* 2009; **14**: 116–20.

14. Lotan M, Moe-Nilssen R, Ljunggren AE, Strand LI. Reliability of the Non-Communicating Adult Pain Checklist (NCAPC), assessed by different groups of health workers. *Res Dev Disabil* 2009; **30**: 735–45.

15. Hauer J, Houtrow AJ; Section on Hospice and Pallitive Medicine, Council on Children with Disabilities. Pain assessment and treatment in children with significant impairment of the central nervous system. *Pediatrics* 2017; **139**: e20171002.

16. Crosta QR, Ward TM, Walker AJ, Peters LM. A review of pain measures for hospitalized children with cognitive impairment. *J Spec Pediatr Nurs* 2014; **19**: 109–18.

17. Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment. *Paediatr Anaesth* 2006; **16**: 258–65.

18. Voepel-Lewis T, Malviya S, Tait AR, et al. A comparison of the clinical utility of pain assessment tools for children with cognitive impairment. *Anesth Analg* 2008; **106**: 72–8.

19. Pedersen LK, Rahbek O, Nikolajsen L, Møller-Madsen B. Assessment of pain in children with cerebral palsy focused n translation and clinical feasibility of the revised FLACC score. *Scan J Pain* 2017; **9**: 49–54.

20. Hidecker MJ, Paneth N, Rosenbaum PL, et al. Developing and validating the Communication Function Classification System for individuals with cerebral palsy. *Dev Med Child Neurol* 2011; **53**: 704–10.

21. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatr Nurs* 1997; **23**: 293–7.

22. Voepel-Lewis T, Merkel S, Tait AR, Trzcinka A, Malviya S. The reliability and validity of the Face, Legs, Activity, Cry, Consolability observational tool as a measure of pain in children with cognitive impairment. *Anesth Analg* 2002; **95**: 1224–9.

23. Raudenbush SW. A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *J Educ Stat* 1993; **18**: 321–49.

24. Samejima F, vandeLiden WJ, Hambleton R. The graded response model. In van der Liden WJ, Hambleton RK, editors. Handbook of modern Item Response Theory. New York: Springer, 1996: 85–100.

Table I: Participant characteristics (*n*=48)

Mean (SD) age	29y 2mo (13y)
Age range (y)	13–64y
Female	35 (73)
Male	13 (27)
Type of CP	
Spastic quadriplegic	31 (65)
Spastic diplegic	9 (19)
Spastic hemiplegic	8 (17)
CFCS level (IV–V not eligible)	
I	29 (60)
II	14 (29)
III	5 (10)
GMFCS level	
I	3 (6)
II	8 (17)
III	10 (20)
IV	26 (54)
V	1 (2)

Data are *n* (%) unless otherwise indicated. CP, cerebral palsy; CFCS, Communication Function Classification System; GMFCS, Gross Motor Function Classification System.

Table II: The revised Face, Legs, Activity, Cry, and Consolability scale with additional individual behaviour comment[17]

Domain	Participant individualized behaviour
Face 0=No particular expression or smile 1=Occasional grimace/frown; withdrawn or disinterested; appears sad or worried 2=Consistent grimace or frown; frequent/constant quivering chin, clenched jaw; distressed-looking face; expression of fright or panic Individualized behaviour:_____	Close eyes Eyes squinting Scratching face Look away Eyes widening Furrowed brow Bite lip Bite arm or hand Grind teeth
Legs 0=Normal position or relaxed; usual tone and motion to limbs 1=Uneasy, restless, tense; occasional tremors 2=Kicking, or legs drawn up; marked increase in spasticity, constant tremors or jerking Individualized behaviour:_____	Spasms Rub and massage thigh Legs shaking Legs jerking Flex legs to chest Legs twist Legs flex
Activity 0=Lying quietly, normal position, moves easily; Regular, rhythmic respirations 1=Squirming, shifting back and forth, tense or guarded movements; mildly agitated (e.g. head back and forth, aggression); shallow, splinting respirations, intermittent sighs. 2=Arched, rigid or jerking; severe agitation; head	Clench fists Rock and lean forward Tilt chair back Laughing Deep breathing Startle Extend arms Flex arms

banging; shivering (not rigors); breath holding, gasping or sharp intake of breaths, severe splinting Individualized behaviour: _____	Snorting Restlessness Arms twitching Withdrawing Walking around Standing very still Crack back Point to location of pain
Cry 0=No cry/verbalization 1=Moans or whimpers; occasional complaint; occasional verbal outburst or grunt 2=Crying steadily, screams or sobs, frequent complaints; repeated outbursts, constant grunting Individualized behaviour: _____	Tears without sound Constant moaning
Consolability 0=Content and relaxed 1=Reassured by occasional touching, hugging or being talked to. Distractible 2=Difficult to console or comfort; pushing away caregiver, resisting care or comfort measures Individualized behaviour: _____	Being quiet in group Music Hugs Touch Hugs self Frustration Hold hands

Table III: Differences in pain scores by rater and participant

	n	%	Mean difference
Rater score lower than participant (range –1 to –3)	52	54	2.91
Rater score higher than participant (range 1–5)	28	29	–1.93

Rater score same as participant	16	17	–
Absolute difference of rater vs participant (0–1 point)	46	48	–
Absolute difference of rater vs participant (2–4 points)	40	42	–
Absolute difference of rater vs participant (5+ points)	10	10	–

Table IV: Rater effects

Effect	Estimate	SE	t	*p*
Lay (vs all else)	0.63	0.73	0.86	0.39
Nurse (vs all else)	0.54	0.80	0.67	0.50
Physician (vs all else)	–0.57	0.54	–1.06	0.29
Inexperienced (vs experienced)	0.72	0.45	1.61	0.11

SE, standard error; t, t-statistic.