Pain 93 (2001) 173–183

PAIN

www.elsevier.nl/locate/pain

The Faces Pain Scale – Revised: toward a common metric in pediatric pain measurement ☆

Carrie L. Hicks[a], Carl L. von Baeyer[a,b,*], Pamela A. Spafford[a], Inez van Korlaar[c], Belinda Goodenough[c]

[a]*Department of Psychology, University of Saskatchewan, Saskatoon, Canada*
[b]*Department of Pediatrics, University of Saskatchewan, Saskatoon, Canada*
[c]*School of Psychology, University of New South Wales, Sydney, Australia*

Received 28 April 2000; received in revised form 26 February 2001; accepted 12 March 2001

Abstract

The Faces Pain Scale (FPS; Bieri et al., Pain 41 (1990) 139) is a self-report measure used to assess the intensity of children's pain. Three studies were carried out to revise the original scale and validate the adapted version. In the first phase, the FPS was revised from its original seven faces to six, while maintaining its desirable psychometric properties, in order to make it compatible in scoring with other self-rating and observational scales which use a common metric (0–5 or 0–10). Using a computer-animated version of the FPS developed by Champion and colleagues (Sydney Animated Facial Expressions Scale), psychophysical methods were applied to identify four faces representing equal intervals between the scale values representing least pain and most pain. In the second phase, children used the new six-face Faces Pain Scale – Revised (FPS-R) to rate the intensity of pain from ear piercing. Its validity is supported by a strong positive correlation ($r = 0.93$, $N = 76$) with a visual analogue scale (VAS) measure in children aged 5–12 years. In the third phase, a clinical sample of pediatric inpatients aged 4–12 years used the FPS-R and a VAS or the colored analogue scale (CAS) to rate pain during hospitalization for surgical and non-surgical painful conditions. The validity of the FPS-R was further supported by strong positive correlations with the VAS ($r = 0.92$, $N = 45$) and the CAS ($r = 0.84$, $N = 45$) in this clinical sample. Most children in all age groups including the youngest were able to use the FPS-R in a manner that was consistent with the other measures. There were no significant differences between the means on the FPS-R and either of the analogue scales. The FPS-R is shown to be appropriate for use in assessment of the intensity of children's acute pain from age 4 or 5 onward. It has the advantage of being suitable for use with the most widely used metric for scoring (0–10), and conforms closely to a linear interval scale.

Keywords: Faces Pain Scale – Revised; Common metric; Pediatric pain measurement

1. Introduction

In the past decade, there has been an increased interest in the assessment and management of pain in pediatric populations. However, according to Kuttner (1996) "attempting to assess and measure another person's pain is like trying to speak a foreign language that you don't understand" (p. 76). Determining the best way to measure pain in children has been difficult, thus far, for two major reasons. First, the assessment of pain in children can present developmentally-specific difficulties due to potentially limited abilities in verbal communication and associative thinking (Champion et al., 1998). Second, although the International Association for the Study of Pain has made recommendations for regular charting of pain, in many settings no such practices have been put in place despite good intentions and the existence of many published scales. As a result of these two relatively independent factors, validated pain measures are not commonly used in pediatric care (Hester et al., 1998).

Without regular pain assessment, pain is often undertreated. Overcoming some of the challenges in assessing children's pain necessitates the development of reliable, valid and age-appropriate pain measures (Tesler et al., 1991). Such tools could be used both with acute pain when measures would be taken several times a day or chronic pain when measures could be taken less often, perhaps once a day. To be most effective, pain measurement tools must be simple, practical, and useful, without impos-

☆ This paper was presented in part as a poster at the Canadian Pain Society 1999 Annual Meeting, St. John's, NF, Canada, 15 May 1999.

* Corresponding author. Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK, S7N 5A5, Canada. Fax: +1-306-966-6630.

E-mail address: carl.vonbaeyer@usask.ca (C.L. von Baeyer).

0304-3959/01/$20.00
PII: S0304-3959(01)00314-1

ing a large burden on caregivers. Simplifying pain measurement may have multiple benefits including: (a) improved relief of pain in children; (b) decreased workload for nurses and other health care professionals related to improved patient pain relief; and (c) creation of a common pain language to facilitate communication about pain within and across settings.

1.1. The need for a common metric

More than 40 published pain measures are available for use with children (Finley and McGrath, 1998). Three types of measures are typically used in health care settings to measure pain: self-report, observational and physiological. Many different metrics are used to assess pain intensity (e.g. 0–5, 0–10, 0–42, and 0–100). In some institutions where different scales are simultaneously in use, because of the lack of a consistent practice of indicating which scale a given score refers to, it is often not possible to tell whether a recorded pain score of, say, 5, indicates severe pain (5/5), moderate pain (5/10) or minimal pain (5/100). As a result there can be significant inconsistency in how and when the measures are applied, if they are even used at all (von Baeyer et al., 1998). One way to eliminate such inconsistency would be the development of a common metric which could be used both within and across institutions. Communication of pain scores in standardized units, much like the intelligence quotient or the Celsius scale, would facilitate shared understanding of the meaning of the scores. A common metric would allow for a conventional and universal language for reporting scores, but would not be limited to a list of particular measurement instruments. It is important to note that the extent to which scores from various pain measures conform to this theoretical linear metric is a question to be answered separately for each tool and for each measurement situation (von Baeyer and Hicks, 2000).

If some consensus can be reached about the most desirable metric to be used, old measures can be adapted or new measures can be developed to fit within this framework. In order to reach a consensus, compromises will need to be made. Most self-report pain intensity measures designed for young children have four to six levels or divisions (e.g. Beyer, 1984; Hester, 1979; Whaley and Wong, 1987). On the other hand, self-report pain intensity measures designed for adults commonly have finer, more numerous divisions (e.g. the NRS-101, a common form of the numerical rating scale that provides a score from 0 to 100). Thus, if a common metric is to be adopted it will represent a compromise between the number of points necessary to measure pain intensity and applicability across the life-span (McCaffery and Pasero, 1999). Different instruments may be needed to accommodate the substantial variability across the pediatric age span, but pain intensity scored on the different instruments can be recorded on the same metric to facilitate communication.

Currently, a movement toward a consensus can be identified in that two metrics are most accepted: 0–5 and 0–10 (McCaffery and Pasero, 1999). Several pre-existing and developing systems in both research and clinical practice have become focused on these two most common metrics. For example, McCaffery and Pasero (1999) suggest a system, to be used throughout a clinical setting, that combines and standardizes a numerical rating scale and the Wong–Baker FACES scale on a 0–10 metric. While some observational and self-report measures already use these metrics, other measures will need to be adapted and validated.

Once measures have a shared metric, they can be more easily combined into measurement systems which employ multiple approaches, but are straightforward and simple to use. Improvement of the assessment and management of pain necessitates more user-friendly measurement tools for health care providers (Hester et al., 1998). Although there are other self-rating scales (e.g. numerical rating scales) that are already based on the 0–5 or 0–10 metric, a psychometrically sound facial expression scale would be an important addition to such a pain measurement system. Face scales are especially good for younger children because these measures are concrete and presumably easier for young children to follow, especially in the 4–8-year-old range (Champion et al., 1998). In order to use this tool, children would need the cognitive capacity to match pictures of facial expressions to their own internal state, but they would not need to be able to count or to use numbers in a categorical fashion. Children also report a preference for the faces scales, in part due to the ease of application (Champion et al., 2000).

1.2. Faces Pain Scale

One candidate for a simplified measurement system is the Faces Pain Scale (FPS; Bieri et al., 1990). It has several advantages over other facial expression scales. In creating this facial scale, Bieri et al. (1990) addressed the problems that had been encountered with other such scales. Specifically, they avoided the problems inherent in inclusion of smiles and tears in a pain scale: (a) the apparent confounding of affective distress with pain intensity (Champion et al., 1998; Chambers and Craig, 1998); and (b) the significantly higher pain ratings given on scales that have smiling ‘no pain’ faces compared with scales that have neutral ‘no pain’ faces (Chambers et al., 1999).

The development and psychometric refinement of the FPS occurred in several stages. The result of this multistage process is a straightforward measure that allows pain intensity to be measured in young children who may have difficulty with more cognitively demanding instruments. The original version of the scale consists of seven faces increasing in pain intensity and approximating equal intervals as assessed by children (Bieri et al., 1990). It must, of course, be acknowledged that a point on such a scale has no

absolute meaning, but is relative to the rater's experience with pain, as well as other factors.

The original version of the FPS would require adaptation as it is a seven-point scale and therefore does not lend itself easily to scaling on either a 0–5 or 0–10 metric. A revision of the scale developed by Bieri et al. (1990) will allow the scores from the FPS to be on the same metric or scale as numerical self-report and observational measures of pain. For the current adaptation of the scale, six faces were considered more advantageous than 11 because of the challenge that the larger number of faces, and consequent finer distinctions, would pose for children aged 3–5 years (Shih and von Baeyer, 1994).

The Sydney Animated Facial Expressions (SAFE) Scale developed by Champion and colleagues (Goodenough et al., 1997) is a version of the FPS which was created in other research to explore the ways in which young children apply scales to estimate pain intensity (i.e. whether or not young children have a bias for the ends of a scale; cf. Chambers and Johnston, 2001). The SAFE scale is a computer animation in which a single face varies smoothly from 'no pain' to 'most pain possible' faces of the FPS (through 101 frames). In Study 1, this animated form of the scale was used in combination with printed copies of the anchor points (i.e. the first and last face as designated by Bieri et al., 1990) and adults were asked to identify the four intermediate faces to be included in the six-face adaptation of the FPS.

1.3. The present methodology and aim

As suggested by McGrath (1987), cross-modality matching (CMM) methodology and related techniques may be ideally suited for psychometric refinement of pain measures. In Study 1, only one modality was used (i.e. the presentation of the pain faces). Two approaches related to CMM, the estimation and production methods, are suggested when examining only one modality. In magnitude estimation, a stimulus is presented and the participant is asked to assign a number to that presentation (Luce and Krumhansl, 1988). Magnitude production reverses the procedure used in magnitude estimation, in that the participant is given a number and asked to produce a matching intensity (Falmagne, 1985). For Study 1, the method of magnitude production was considered most appropriate for this investigation, as it allowed participants to match the pain intensity represented by a face to a number. Using this technique facilitated the reduction of the number of faces in the scale, while minimizing the number of trials required to obtain reliable results. Additionally, this method was more analogous to the way the scale will later be used in clinical practice (i.e. users will be asked to select a face matching the intensity of their pain experience).

In Studies 2 and 3, the CMM method was used with children matching their ratings of pain using pain faces and line length or color gradation (i.e. the visual analogue scale (VAS) and colored analogue scale (CAS), respectively). In CMM, the task for participants is to equate the perceived strength of two stimuli from two different modalities (Stevens, 1960). An experimenter selects a series of stimuli (e.g. sounds) and measures each by physical means (e.g. decibel meter). The participant might then adjust the intensity of a light to match each presented intensity of sound and the experimenter measures the responses (e.g. photometer). Such techniques have been used successfully even with children as young as age 4 because they do not depend on the understanding of numbers or words (McGrath, 1990).

The first aim of the present research was to revise the FPS from seven faces to six faces to make it practical to score it either 0-1-2-3-4-5 or 0-2-4-6-8-10, thereby allowing the FPS to be incorporated into a simplified pain assessment system that utilizes both self-report and observational measures employing the 0–5 or the 0–10 metric. The second aim was to evaluate the validity of the revised six-face version (Faces Pain Scale – Revised, FPS-R) by using the scale with children in actual pain situations, in both non-clinical and clinical settings. Ratings obtained on the revised scale were compared with those on the VAS which has been well-documented as a reliable and valid pain measurement tool (McGrath, 1987), and the CAS (McGrath et al., 1996).

2. Study 1: derivation of the FPS-R from the FPS

2.1. Method

2.1.1. Participants

Following the methods of Gracely et al. (1978) (see Section 2.2), a sample of 15 adults was obtained from the student population at the University of Saskatchewan. Participation was voluntary and some individuals received course credit for participating. Adults were selected for this phase because the task was considered too complex and abstract for children; however, all participants had extensive experience with children between the ages of 5 and 12. Parents, and others who had extensive exposure to children, were specifically selected because their familiarity with and demonstrated interest in children was seen as beneficial and likely to increase the seriousness with which they would approach the task. They were informed that they were assisting in the development of a scale that would be used, in future, to measure pain in children. Participants ranged in age from 18 to 44 and included parents, coaches, swimming instructors, nannies and other caregivers, and an elementary school teacher.

2.1.2. Materials

The SAFE scale is a continuous computerized presentation of the FPS consisting of 101 frames with very subtle graphical differences between each frame. When presented at a speed of approximately 11 frames per second, the resulting animation is a single face changing smoothly from 'no

pain' to 'very much pain' or vice versa. Participants were not told that they were selecting faces from a 101 frame series. The SAFE scale was presented to participants on an IBM compatible PC via a software media player, with the animation controlled by depression of cursor keys on the keyboard. Pressing the left arrow key caused the face to change in the direction of less pain, and the converse for the right arrow key. Paper copies of the no pain and very much pain expressions, assigned the numbers 0 and 5, respectively, were provided and used for reference.

2.1.3. Procedure

Following institutional ethics approval, participants were recruited and informed consent was obtained. Participants were tested individually in a single session lasting approximately 45 min. Seated in front of a computer which displayed the SAFE scale, each participant was asked to scroll back and forth through the facial expression so that they (a) became comfortable with how to change the face, and (b) identified the range and endpoints of the scale.

Using the magnitude production method, on each trial the participant was told by the experimenter a number between the 0 and 5 endpoints (i.e. 1, 2, 3, or 4) and asked to adjust the facial expression until it was perceived to correspond to that scale value of pain intensity. The numbers were given in one of five randomly generated orders. On each trial, the researcher recorded the frame number corresponding with the facial expression selected (a score of 0–100 inclusive), without the participant seeing the score. There were 20 trials per stimulus (i.e. the numbers 1, 2, 3, and 4) for a total of 80 trials per participant. After each trial the animation was reset by the researcher to either frame 0 or frame 100 in a counterbalanced order. The experiment began with a practice phase comprising four trials that were not included in the results analysis (i.e. one for each of the numbers 1–4).

2.2. Results and discussion

Reflecting the repeated measures design, the data set consisted of 1200 data points (300 ratings for each of the four stimulus values). Mean and median rating values were calculated for each participant, for each of the four stimuli, yielding very similar values. The means (across participants) associated with scale values 1, 2, 3, and 4 were found to be 25, 44, 62 and 80, respectively. The faces corresponding with these frame numbers are represented in the FPS-R shown in Fig. 1, below the original version of the FPS.

The standard error of the mean (SEM) was calculated for each stimulus level. As Fig. 2 demonstrates, the differences in facial representations within 1 SEM were extremely subtle, providing greater confidence in the validity of the particular face chosen at each scale value. For three of the stimuli, the differences were almost impossible to discern; however, for stimulus value 1, the differences were somewhat more easily detected due to the presence/absence of a single graphical feature (i.e. small forehead furrow).

Further support for the validity of the face selected at each of the hypothetical pain intensity values is demonstrated in Fig. 3. Fig. 3 depicts the linear and non-overlapping distributions of SAFE faces selected across the four stimulus levels.

Intra-class correlation was also calculated to assess inter-rater agreement for the entire stimulus set. Calculation of intra-class correlation allows for an evaluation of agreement between raters when the measurements are interval, with the resulting coefficient indicating degree of concordance (Portney and Watkins, 1993). Using a two-way mixed effect

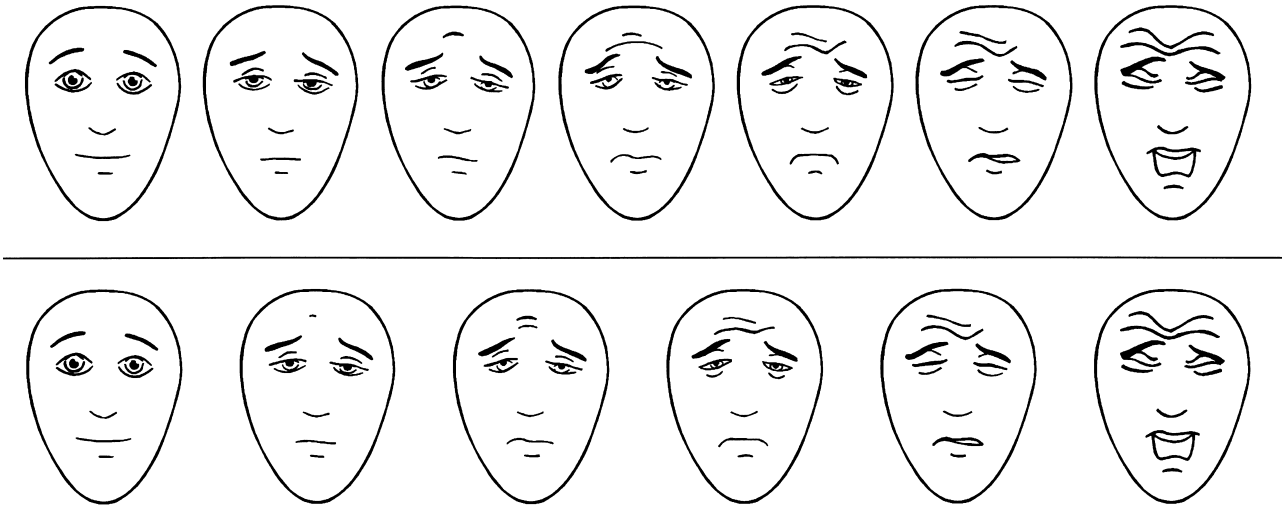

Fig. 1. Top: Faces Pain Scale (Bieri et al., 1990), scored 0 to 6. Bottom: Faces Pain Scale-Revised, scored 0-2-4-6-8-10 (or 0-1-2-3-4-5). Instructions: "These faces show how much something can hurt. This face [point to left-most face] shows no pain. The faces show more and more pain [point to each from left to right] up to this one [point to right-most face] - it shows very much pain. Point to the face that shows how much you hurt [right now]".

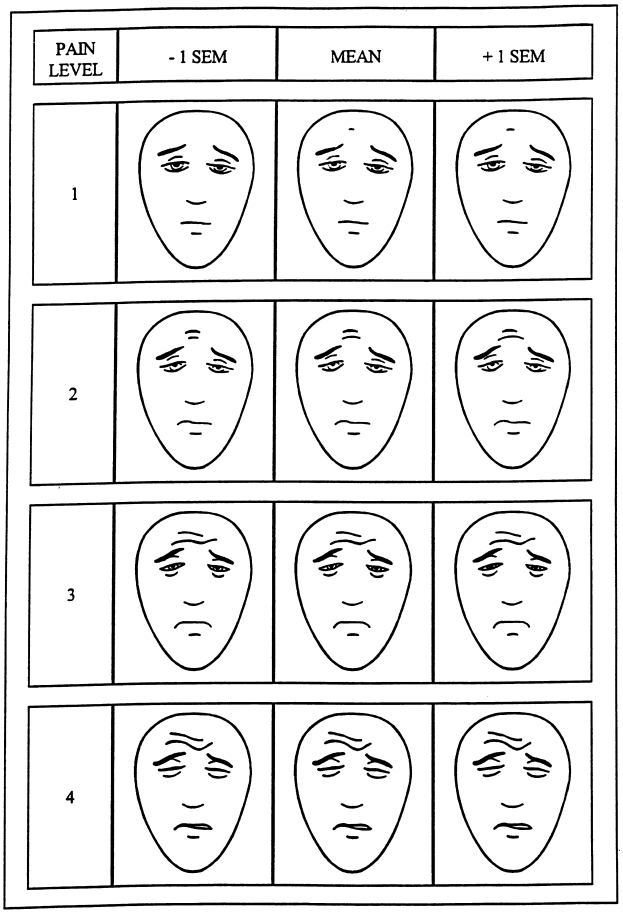

Fig. 2. Study 1: mean face selected (±1 SEM) by hypothetical pain intensity level for FPS-R.

model, the single measure intra-class correlation coefficient was 0.94, representing a high degree of concordance among the raters. Inter-rater correlations, calculated for each pair of raters, ranged from 0.84 to 0.99.

Rater agreement was further assessed using a method whereby the ratings of each rater were compared with those of the overall group. Strong positive correlations were found between the ratings from each participant and the overall group ratings (mean $r = 0.97$, range $r = 0.91$–0.99, all $P < 0.001$). These strong positive correlations further suggest high levels of rater agreement, allowing users to have greater confidence with respect to the validity of the face chosen at each stimulus value (Uebersax, 1988, 2000).

To further demonstrate the psychometric qualities of the revised scale, a between-group comparison and post-hoc analysis were performed using a method adapted from Gracely et al. (1978). The fifteen participants were divided into two groups of five and ten individuals, respectively, with the split based on consecutive participation. This asymmetry in group size allowed for a greater number of between-group comparisons than would have been possible with a more symmetrical division, because some analyses involved correlating individual ratings (of each individual from the larger group) with the mean of the smaller group. No significant group differences in mean (of individual median) ratings were found between Group 1 and Group 2 for any of the four stimulus values (see Table 1).

Further analyses were performed to test the relationship between the ratings of participants in Group 1 and those in Group 2. Using a method adapted from Gracely et al. (1978), with a comparable sample size, the ratings from each participant in Group 2 were compared to the mean Group 1 rating. Strong positive correlations were found between the ratings from each of the ten Group 2 members and the mean Group 1 rating (mean $r = 0.94$, range $r = 0.91$–0.99, all $P < 0.001$).

3. Study 2: validation of the FPS-R with a painful procedure

Study 1 demonstrated the psychometric qualities of the six-face FPS-R by establishing the linear progression of the faces and by presenting several comparisons (both within and between groups) based on 1200 ratings by adults. The next step in the development of the FPS-R was to determine the validity of the measure with children who are experiencing 'real-life' pain, namely ear piercing. The convergent validity of the FPS-R was assessed by correlating it with a VAS. As the age range selected for this phase was 5–12 years, possible age effects were also examined.

3.1. Method

3.1.1. Participants

The sample consisted of 76 children requesting ear piercing at two urban jewelry stores. The present data were obtained in conjunction with another study (Spafford et al., 2001). Children who assented to participate in the study were included only if they (a) ranged in age from 5

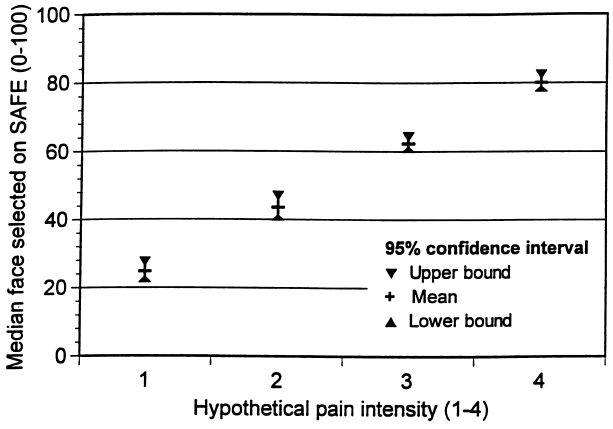

Fig. 3. Study 1: median face selected by hypothetical pain intensity level showing 95% confidence interval.

Table 1
Study 1: descriptive statistics and comparison of group means by stimulus: analysis based on the method of Gracely et al. (1978)

Stimulus	Group 1 ($N = 5$)		Group 2 ($N = 10$)		ANOVA ($F_{(1,13)}$, P)
	Mean[a]	SD	Mean[a]	SD	
1	25.6	5.22	24.7	6.36	0.08, 0.78
2	45.8	3.25	42.8	8.24	0.62, 0.45
3	63.4	4.04	62.2	4.61	0.24, 0.63
4	78.3	4.84	81.3	4.86	1.23, 0.29

[a] Frame number (0–100) where each number represents one face on a continuum from no pain to very much pain.

to 12 years, (b) were accompanied by a parent who provided consent for their participation, and (c) were having their ear(s) pierced for the first time. Two parents declined consent for their child to participate. The age range from 5 to 12 years was selected to address concerns due to the measuring instruments (i.e. children under the age of 5 tend to use the extreme ends of the scale; Arts et al., 1994) and the experimental situation (i.e. children younger than 5 may have more difficulty in coping with the pain and at this age there is a greater likelihood that the ear piercing is not voluntary). Children over the age of 12 were not included because it has been reported that it is likely that most children of this age do not experience significant distress during ear piercing (von Baeyer et al., 1997).

The sample included 57 females and 19 males who ranged in age from 5 to 12 years, with a mean age of 7.67 years. The sample was divided into three age groups (Table 2).

3.1.2. Materials

3.1.2.1. VAS. The VAS was selected as it has been consistently described as a well-validated measure of pain (e.g. McGrath, 1987). The version used here was a 200 mm scale that consisted of a single horizontal black line, with endpoints labeled 'no pain' and 'very much pain'. Each child was asked to put a mark on the line which corresponded with his or her amount of hurt or pain. The VAS was scored 0–100, with 1 unit equivalent to 2 mm.

3.1.2.2. FPS-R. The FPS-R is presented in the bottom panel of Fig. 1. Each child was asked to mark the face that reflected his or her experienced pain. Again, the endpoints were labeled 'no pain' and 'very much pain'. The FPS-R was scored 0–5.

Table 2
Study 2: means, standard deviations (SD), and correlations between VAS and FPS-R ratings by age group

Age group (years)	N	VAS (0–100)		FPS-R (0–5)		$r_{(VAS, FPS-R)}$[a]
		Mean	SD	Mean	SD	
5–6	24	46.2	35.2	2.25	1.96	0.94
7–8	27	35.3	29.7	1.93	1.64	0.92
9–12	25	30.9	27.1	1.76	1.39	0.93
Total	76	37.3	31	1.97	1.67	0.93

[a] All $P < 0.001$.

3.1.3. Procedure

Following institutional ethics approval, participants were recruited and informed consent was obtained from parents with assent obtained from their children. The data were collected in the context of a study on the effect of parental information on children's expectations of pain during ear piercing. Participants were not recruited until after they had made the decision to have the ear piercing done, so that this decision was not influenced by the research. The children requesting ear piercing and meeting the inclusion criteria were approached to participate and offered a $4 discount off the price of ear piercing. Each participant was tested individually. Store staff tested 64 participants following training and demonstration by the third author who herself tested 12 participants. Before the ear piercing, each participant had two practice trials with the FPS-R and the VAS (two per scale), using imagined situations which helped them to become familiar with the measures. The children were given a no pain example (chewing bubble gum) and a pain example that would likely have been experienced by most participants (getting a needle) and asked to rate their pain in these situations. Also, each participant had four additional trials with the FPS-R and the VAS (two per scale), measuring expected pain from ear piercing. (These expectancy measures, included for another study, were not included in the present analyses.) Piercing was carried out simultaneously for girls (i.e. both ears were done at the same time) while all the boys elected to have only one ear pierced. Using both the VAS and the FPS-R, with the order of presentation counterbalanced across participants, the children were asked to rate how much hurt they had experienced.

3.2. Results and discussion

A strong positive correlation ($r = 0.93$, $P < 0.001$) was found between the ratings of pain intensity on the VAS and FPS-R. Fig. 4 demonstrates this relationship. Correlations between the ratings on the VAS and the FPS-R were also computed with participants divided into age groups (see Table 2). A strong positive correlation was found both over-

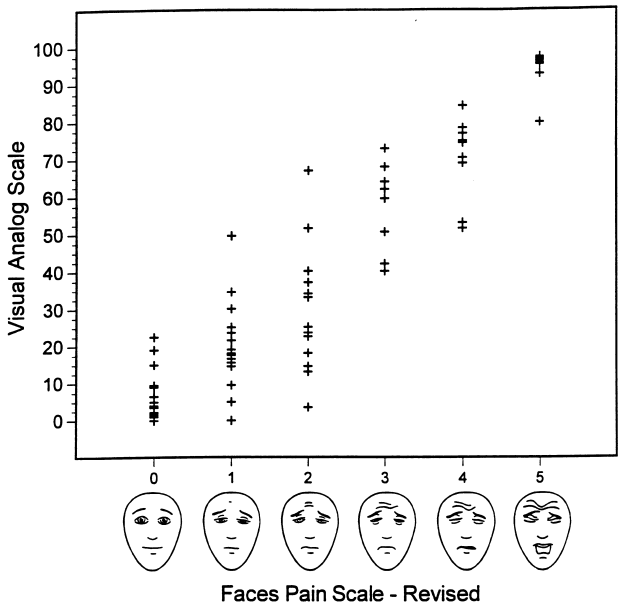

Fig. 4. Study 2: scatter plot of the relationship between FPS-R and VAS ratings of ear piercing pain in children aged 4–12 ($r = 0.93$, $N = 76$).

all and for each of the three age groups (5–6, 7–8, and 9–12 years).

The FPS-R ratings of males (mean 1.7, SD 1.5) and females (mean 2.1, SD 1.7) were not significantly different ($F_{(1,74)} = 0.51$, $P = 0.48$), although it may be of interest that none of the males but seven of the females rated their pain using the highest pain intensity face (i.e. 5). (Note, however, that all the girls had two ears pierced while the boys had one ear pierced.) Similarly, the VAS ratings of males (mean 30.8, SD 27.2) and females (mean 39.4, SD 32.1) were not significantly different ($F_{(1,74)} = 1.12$, $P = 0.29$). Pain ratings were not significantly correlated with age on either FPS-R ratings ($r = -0.19$, $P = 0.10$) or VAS ratings ($r = -0.12$, $P = 0.32$). However, none of the 10-, 11- or 12-year old participants rated their pain higher than a 3 on the FPS-R, while 33.3% of the 5- and 6-year-olds did so.

The strong positive correlation between the pain intensity ratings on the VAS and the FPS-R provides evidence for the concurrent validity of the new six-face instrument. Thus, the present study suggests that the FPS-R is a valid self-report measure of the pain intensity perceived by a child in an actual pain situation; however, confidence in the validity of the FPS-R is supported here only within the age range of 5–12 years.

Numerous previous reports (reviewed by Champion et al., 1998) suggest that some 5-year-old children may have difficulty using self-report scales such as the VAS and the FPS-R. Interestingly, the current study demonstrated that many 5-year-old children may, in fact, be able to use such measures in a reliable manner, similar to older age groups. Further investigation with a larger group is suggested to explore the utility of such self-report measures with children under 6 years.

Table 3 demonstrates a progressively increasing relationship between the VAS and FPS-R for each of the age groups. Despite the lack of age main effects, there do appear to be some age interactions. One important feature is the distribution of responses: none of the 10–12-year-olds selected high pain scores (i.e. above a 3), but the full range of pain scores from face 0 to face 5 was used by participants younger than 10 years. This finding attests to the validity of the ear piercing paradigm as a useful non-clinical analogue of needle-related procedural pain (cf. Arts et al., 1994).

In the current study, three important considerations led to the choice of ear piercing pain as the real-life painful stimulus: (a) exposure to this painful stimulus is voluntary and is not initiated by the researcher; (b) the procedure is brief; and (c) the equipment and procedure are standard. Given that the FPS-R has been validated on this non-health care population, consideration must be given to the qualitative aspect of the pain. The pain experienced during ear piercing, which is a procedure usually desired and/or requested by the child, is probably different from that experienced in clinical situations. Accordingly, further validation studies should be done using the FPS-R in other populations, such as those experiencing procedural pain (e.g. venipuncture) or chronic pain (e.g. arthritis).

4. Study 3: validation of the FPS-R with a clinical sample

Study 2 demonstrated the validity of the FPS-R with children who are experiencing 'real-life' but non-clinical pain. The next step in the development of the FPS-R was to determine the validity of the measure with children experiencing clinical pain. The convergent validity of the FPS-R was assessed by correlating it with the VAS and CAS. As the age range selected for this phase was 4–12 years, possible age effects were also examined.

4.1. Method

4.1.1. Participants

The participants were a convenience sample of 90 inpatients consecutively contacted at a children's hospital to

Table 3
Study 2: mean VAS score by FPS-R score and age group

Age group (years)	FPS-R score		
	0–1	2–3	4–5
5–6	14.0 (10)[a]	45.2 (6)	87.1 (8)
7–8	13.0 (13)	39.6 (7)	72.3 (7)
9–12	11.0 (12)	39.0 (10)	83.0 (3)
Total	12.6 (35)	40.8 (23)	80.7 (18)

[a] Numbers in parentheses represent the number of children.

participate in a study of pain measurement. Inclusion criteria specified that children were contacted for participation only if they (a) were 4–12 years old, (b) were accompanied by a parent who gave consent for the study, and (c) were experiencing pain either at the time of testing or since their admission to hospital. There were 47 males and 43 females with a mean age of 8.6 years (SD 2.6). The children were divided for analysis into three equal age groups, each with 30 children: 4–6 years, 7–9 years, and 10–12 years.

Reasons for hospitalization were for surgical treatment in 68 cases (75%) including abdominal (18), ear/nose/throat (12), orthopedic (12), urological (7), and other (19). The remaining 22 cases (25%) were hospitalized for non-surgical painful conditions: abdominal (5), respiratory (5), orthopedic/rheumatological (4), and other (8).

4.1.2. Materials

4.1.2.1. FPS-R. The FPS-R consisted of a card showing the six facial expressions determined in Study 1 (see the bottom panel of Fig. 1). Each face was 25×35 mm with 13 mm between faces. Each child was asked to point to the face that reflected his or her current or recalled pain (see Section 4.1.3). The endpoints were explained as 'no pain' and 'very much pain'. The FPS-R was scored 0–10.

4.1.2.2. VAS. The VAS consisted of a mechanical device with a plastic slider showing a horizontal 10 cm black line with ticks 1 cm apart (Abu-Saad et al., 1994). Each child assigned to use the VAS was asked to move the slider to the position that reflected his or her current or recalled pain. The VAS was scored 0–10 in increments of 0.25. The endpoints were labeled 'no pain' and 'most pain'.

4.1.2.3. CAS. The CAS consisted of a mechanical device with a plastic slider over a 143 mm long tetragon varying from narrow (10 mm) and white at the end labeled 'no pain' to wide (30 mm) and dark red at the end labeled 'most pain', as described by McGrath et al. (1996). The CAS was scored 0–10 in increments of 0.25.

4.1.3. Procedure

After informed consent was obtained from the parent and child, the child was asked to estimate his or her current pain on the FPS-R, followed by either the VAS or the CAS, selected on a pre-determined random schedule. If the child was in no pain at the time, then he or she was asked to rate the worst pain recalled since admission to hospital. Thus, no ratings of 0 were obtained on the FPS-R in this study. Current pain was rated by 40 children and recalled worst pain was rated by 50 children.

Each child was randomly assigned to use either the VAS or the CAS. The CAS was used by 45 children, and the VAS was used by the other 45 children.

Table 4
Study 3: means, standard deviations (SD), and correlations between VAS and FPS-R ratings by age group ($N = 45$)

Age group (years)	N	FPS-R (0–10)		VAS (0–10)		$r_{(\text{VAS, FPS-R})}$[a]
		Mean	SD	Mean	SD	
4–6	17	6.6	3.2	6.7	3.1	0.93
7–9	14	4.3	2.3	3.7	2.4	0.87
10–12	14	5.3	2.8	5.2	3.3	0.90
Total	45	5.5	2.9	5.3	3.1	0.92

[a] All $P < 0.001$.

4.2. Results and discussion

Pearson correlations between the FPS-R and the CAS and between the FPS-R and the VAS for the three age groups are presented in Tables 4 and 5, and a scatter plot collapsing across age groups and forms of analogue scale is shown in Fig. 5. The correlations between the FPS-R and both of the analogue scales exceeded a common conventional standard ($r > 0.70$) for all three age groups.

As seen in Tables 4 and 5, mean scores on the FPS-R were similar to those on the analogue scales. Comparisons of the means using paired t-tests were non-significant for comparisons at the three age groups (4–6 years, $t_{29} = 1.58$, $P = 0.12$; 7–9 years, $t_{29} = 0.74$, $P = 0.47$; 10–12 years, $t_{29} = 0.26$, $P = 0.80$).

Correlations between the FPS-R and analogue scales were computed separately by year for the 22 children in the youngest age group. These were as follows: 4-year-olds, $r = 0.86$, $N = 9$; 5-year-olds, $r = 0.84$, $N = 11$; 6-year-olds, $r = 0.94$, $N = 10$. Even among the youngest children in the sample there was evidence of usage of the FPS-R and analogue scales in a consistent and reliable manner.

The correlations for the youngest age group might have been inflated by the well-documented tendency for young children to select the extremes on scales (Chambers and Johnston, 2001). To the extent that children selected a score of 10 on both the FPS-R and the analogue scales as a matter of response bias rather than a considered self-assessment of pain severity, the correlation would give an inflated estimate of the concurrent validity. (The proportions of chil-

Table 5
Study 3: means, standard deviations (SD), and correlations between CAS and FPS-R ratings by age group ($N = 45$)[a]

Age group (years)	N	FPS-R (0–10)		CAS (0–10)		$r_{(\text{CAS, FPS-R})}$[b]
		Mean	SD	Mean	SD	
4–6	13	6.9	3.6	5.7	3.5	0.84
7–9	16	4.5	2.5	4.6	2.7	0.80
10–12	16	4.0	2.5	4.2	2.5	0.91
Total	45	5.0	3.1	4.8	2.9	0.84

[a] CAS and VAS (Table 4) were collected from separate samples.
[b] All $P < 0.001$.

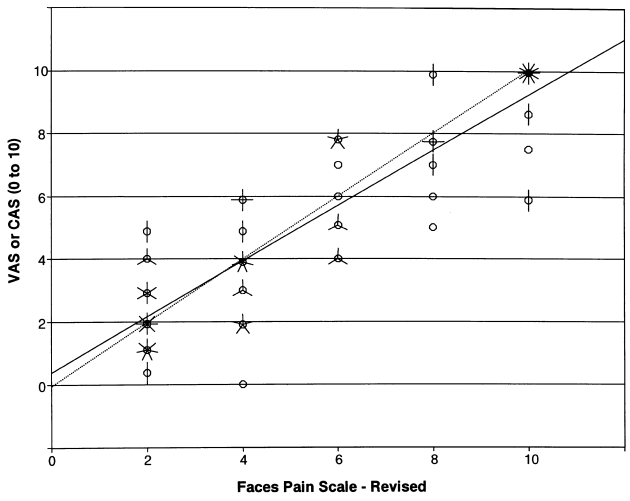

Fig. 5. Study 3: sunflower plot of the relationship between FPS-R and analogue scale ratings of clinical pain in hospitalized children aged 4–12 ($r = 0.88$, $N = 90$). CAS ($N = 45$) and VAS ($N = 45$) scores are combined. Each 'petal' represents one case. The solid line shows linear regression best fit; the dashed line shows the ideal relationship between the FPS-R and analogue scales.

dren who rated their pain at 10 on both the FPS-R and either analogue scale were 27, 3, and 3% in the youngest, middle and oldest age groups, respectively.) In order to test for this effect, an analysis was carried out dropping all children in the youngest age group who scored 10 on both the FPS-R and either analogue scale. The correlation between the FPS-R and the analogue scales remained strong ($r = 0.79$, $N = 22$) (in comparison with $r = 0.87$, $N = 30$ when all children in the youngest age group were included), indicating that young children's tendency to select extremes did not have a major contaminating effect on these concurrent validity estimates.

Children who were in no pain at the time of data collection were asked to rate instead their recalled worst pain during the current hospitalization. Correlations between the FPS-R and the analogue scales were similar for current ($r = 0.86$, $N = 40$) and recalled ($r = 0.84$, $N = 50$) pain. Recalled worst pain was significantly more severe (FPS-R: mean 6.6) than current pain (mean 3.6; $t_{88} = 5.3$, $P < 0.001$).

The original FPS (Bieri et al., 1990) has been extensively validated in clinical studies, and has consistently shown strong correlations with other measures of pain such as the VAS (e.g. Goodenough et al., 1999). The present data show that the FPS-R, like the FPS, exceeds conventional requirements for validity of research tools (e.g. $r > 0.70$) in a sample of pediatric patients selected for pain who used the full range of the scales to describe their pain intensity. The results are consistent with those of Study 2 in showing excellent inter-scale agreement even in the youngest age group, which included children as young as 4 in the present study.

5. General discussion

The goal of the present research was to contribute to (a) the improvement of pain measurement methods used in pediatric populations and (b) the development of a common metric. In Study 1, a six-face adaptation of the FPS was developed for potential inclusion in a simplified system of pain assessment based on a 0–5 or 0–10 metric.

There are several reasons for choosing a face scale, and specifically the FPS, for adaptation to a common metric. Face scales are generally preferred, especially by younger children, to visual analogue, word descriptor, and other types of scales (e.g. Champion et al., 2000). Among the various face scales, the FPS (Bieri et al., 1990) has the advantages of having no smiling face and no tears, thus avoiding the confounding of affect and pain intensity and other complicating effects of these anchors (Chambers and Craig, 1998; Champion et al., 1998).

The notion of "potential provider burden" (Hester et al., 1998, p. 194) appears to be an issue both in using measures routinely and/or introducing new systems. According to Hester et al. (1998), nurses who are typically on the 'front lines' in treating children's pain have concerns about the amount of time that a measurement tool may add to their schedules. Another concern for nurses is that measuring pain routinely is not easy to do on every child, and even if it is done, there is no guarantee that other nurses treating the same child will be consistent in their measurement. Finally, Hester et al. (1998) describe a lack of enthusiasm by some health care professionals with regard to self-report tools; they would rather use their own judgments without asking the child.

These concerns of health care professionals may be allayed to some extent by the adoption of a common metric. The simplified system suggested by von Baeyer et al. (1998) is one example of applying the common metric in practice: it incorporates a numerical self-rating scale for older children and adolescents, the FPS-R for younger children, and behaviorally anchored global rating scales for children unable to provide a meaningful self-report; all are based on a 0–10 common metric. Measures based on a common metric can be designed to be time-effective and simple to use and to facilitate consistency between users. According to Hester et al. (1998), when introducing new systems or methods, the "simplicity of the innovation is essential for more rapid adoption" (p. 194).

Despite the potential simplicity of the suggested simplified system (and similar systems) and the incorporation of a common metric, clinical adoption of such methods is a difficult process requiring both health care professional and organizational acceptance (Hester et al., 1998). In part, it would require endorsement of the proposition that by adopting such a system, nurses and other health care professionals may, in the long run, reduce rather than increase their workload. If pain is monitored systematically and efficiently, and therefore managed more adequately,

then one could imagine that overworked nurses would have to deal with less distress and fewer demands. Treating pain as a 'fifth vital sign' for regular charting has been suggested as a way to facilitate the adoption of pain measurement systems, with an emphasis on ameliorating provider burden (World Health Organization, 1998).

The main benefit of developing the FPS-R is that the measure can now be used in either a 0–5 or a 0–10 system, both of which are candidates for the 'common metric'. As McGrath (1999) put it, alluding to the notion of number bias, "0-to-6 does not make sense to us, but 0-to-5 does." According to Baird and Noma (1978), people have systematic biases with numbers which can subtly affect behavior, and multiples of 5 and 10 are clearly favored. By adapting the FPS to a metric that is clearly favored and is used in other pain scales, there is the potential of increased use of the scale, as people prefer to use things that 'make sense'. A recent survey of health care providers subscribing to a pediatric pain internet mailing list showed strong support for use of a common metric in pediatric pain assessment. A 0–10 scale was strongly favored over all alternatives (0–4, 0–5, 0–100, and others) as the standard metric to be used for various ages, instruments, and sources of pain report (von Baeyer and Hicks, 2000). Given that the 0–10 metric is favored, it is important to note that the FPS-R can be recorded as 0–10, as demonstrated in Study 3, with no loss of its desirable psychometric properties.

Young children often encounter difficulty in using self-report measures. However, checks can be done to determine the suitability of a measure for a particular child. McGrath (1990) suggested a method for checking an individual child's capacity to use the VAS in a valid fashion. In this method, referred to as calibration, a correlation is computed between the child's VAS ratings of the diameters of printed circles and the actual diameters; if the correlation exceeds a criterion (e.g. $r > 0.70$) then the child is considered to be able to use the VAS reliably. Similarly, one can carry out an informal check of whether a particular young child (i.e. 5 years and under) is able to use the FPS-R by asking for ratings of pain in three to five hypothetical situations representing a range from no pain to severe pain. A criterion of capacity to use the instrument would be reached if the FPS-R scores correspond to the predicted order of stimulus intensity for the respective hypothetical situations.

Several important directions for future study are highlighted in the present work. An important next step in the validation of the FPS-R is applying it more widely with children who are experiencing procedural and disease-related pain. Investigations are also warranted to establish 'calibration' methods to determine whether individual children are able to use the FPS-R in a valid and consistent manner. The words used to explain the scale, particularly the top anchor (e.g. very much pain vs. most pain possible), may have an important influence on how the scale is understood and applied by children, and this requires experimental study. Finally, an important direction for future research will be assessing the improved clinical outcomes due to implementation of regular pain assessment using the FPS-R and other methods. In the long term, these research directions will lead to improved assessment and management of pain in pediatric populations.

Acknowledgements

The authors would like to acknowledge the contributions of Duc Trieu (Faculty of Medicine, University of New South Wales), who was responsible for developing the 101 faces (from the original seven of the FPS) for the SAFE scale, and Robert Kovacic, who did the programming and Windows 95 applications. The contributions of Louise Alexitch, Tammy Marche, Kellie Hadden and Stephen Shaw are acknowledged. We would also like to express our appreciation to the staff of Ardene's for their commitment to, and assistance with, this project.

References

Abu-Saad HH, Pool H, Tulkens B. Further validity testing of the Abu-Saad Paediatric Pain Assessment Tool. J Adv Nurs 1994;19:1063–1071.

Arts SE, Abu-Saad HH, Champion GD, Crawford MR, Fisher RJ, Juniper KH, Ziegler JB. Age-related response to lidocaine-prilocaine (EMLA) emulsion and effect of music distraction on the pain of intravenous cannulation. Pediatrics 1994;93:797–801.

Baird J, Noma E. Fundamentals of scaling and psychophysics. New York: Wiley, 1978.

Beyer JE. The Oucher: a user's manual and technical report. Evanston, IL: Hospital Play Equipment, 1984.

Bieri D, Reeve R, Champion G, Addicoat L, Ziegler J. The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation and preliminary investigation for ratio scale properties. Pain 1990;41:139–150.

Chambers C, Craig KD. An intrusive impact of anchors in children's faces pain scales. Pain 1998;78:27–37.

Chambers C, Johnston C. Developmental differences in children's use of pain rating scales. J Pediatr Psychol 2001 in press.

Chambers C, Giesbrecht K, Craig KD, Bennett S, Huntsman E. A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. Pain 1999;83:25–35.

Champion GD, Goodenough B, von Baeyer CL, Thomas W. Measurement of pain by self-report. In: Finley GA, McGrath PJ, editors. Measurement of pain in infants and children, Seattle, WA: IASP Press, 1998, pp. 123–160.

Champion GD, Goodenough B, Wu E, Chua K, Taplin JE, Ziegler JB. Children's self-report of pain intensity using six different scale types. Paper presented at the Seventh International Symposium on Pediatric Pain, June 2000, London.

Falmagne J. Elements of psychophysical theory. Oxford: Clarendon Press, 1985.

Finley GA, McGrath PJ. Measurement of pain in infants and children. Seattle, WA: IASP Press, 1998.

Goodenough B, Champion GD, von Baeyer C, Ziegler JB. Facial expression scales for self-report of pain by children: rationale, validity and utility. Paper presented at the Fourth International Symposium on Pediatric Pain, July 1997, Helsinki.

Goodenough B, Thomas W, Champion GD, Perrott D, Taplin JE, von Baeyer CL, Ziegler JB. Unravelling age effects and sex differences in

needle pain: ratings of sensory intensity and unpleasantness of venipuncture pain by children and their parents. Pain 1999;80:179–190.
Gracely RH, McGrath PA, Dubner R. Ratio scales of sensory and affective verbal pain descriptors. Pain 1978;5:5–18.
Hester NK. The preoperational child's reaction to immunization. Nurs Res 1979;28:250–255.
Hester N, Foster R, Jordan-Marsh M, Ely E, Vojir C, Miller K. Putting pain measurement into practice. In: Finley GA, McGrath PJ, editors. Measurement of pain in infants and children, Seattle, WA: IASP Press, 1998, pp. 179–198.
Kuttner L. A child in pain: how to help. what to do. Vancouver, BC: Hartley & Marks, 1996.
Luce RD, Krumhansl CL. Measurement, scaling and psychophysics. In: Atkinson RC, Herrenstein RJ, Lindzey G, Luce RD, editors. Stevens' handbook of experimental psychology, New York: Wiley, 1988, pp. 3–74.
McCaffery M, Pasero C. Pain: clinical manual. St. Louis, MO: Mosby, 1999.
McGrath PA. An assessment of children's pain: a review of behavioral, physiological and direct scaling techniques. Pain 1987;31:147–176.
McGrath PA. Pain in children: nature, assessment and treatment. New York: Guilford Press, 1990.
McGrath PJ. New developments in pediatric pain assessment. Paper presented at the annual convention of the Canadian Psychological Association, May 1999, Halifax, NS.
McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain 1996;64:435–443.
Portney LG, Watkins MP. Foundations of clinical research: applications and practice. Norwalk, CT: Appleton & Lange, 1993.
Shih AR, von Baeyer CL. Preschool children's seriation of pain faces and happy faces in the Affective Facial Scale. Psychol Rep 1994;74:659–665.
Spafford P, von Baeyer CL, Hicks CL. Expected and reported pain in children undergoing ear piercing: a randomized trial of preparation by parents. Behav Res Ther 2001 in press.
Stevens S. Ratio scales, partition scales and confusion scales. In: Gulliksen H, Messick S, editors. Psychological scaling: theory and applications, New York: Wiley, 1960. pp. 49–66.
Tesler M, Savedra M, Holzemer W, Wilkie D, Ward J, Paul S. The word-graphic rating scale as a measure of children's and adolescents' pain intensity. Res Nurs Health 1991;14:361–371.
Uebersax J. Validity inferences from interobserver agreement. Psychol Bull 1988;104:405–416.
Uebersax J. Agreement on interval-level ratings, 2000. Available on-line at http://ourworld.compuserve.com/homepages/jsuebersax/cont.htm.
von Baeyer CL, Hicks CL. Support for a common metric for pediatric pain intensity scales. Pain Res Manage 2000;5:157–160.
von Baeyer CL, Carlson G, Webb L. Underprediction of pain in children undergoing ear piercing. Behav Res Ther 1997;35:399–404.
von Baeyer CL, Barry J, Oliver D, Alberts L, Kosteniuk L. Introduction of a simple standard system for reporting pain in pediatric wards (abstract). Pain Res Manage 1998;3:58.
Whaley L, Wong DL. Nursing care of infants and children. 3rd ed. St. Louis, MO: Mosby, 1987.
World Health Organization. Cancer pain relief and palliative care in children. Geneva: World Health Organization, 1998.