Cross-Validation of a Self-Report Scale for Postoperative Pain in School-Aged Children

Suwannee Suraseranivongse, MD*, Thunyanit Montapaneewat, Dip Nursing and Midwife*, Jitprapa Manon, BN*, Petcharee Chainchop, BSc (Nursing and Midwife)*, Suppat Petcharatana, BN*, Kanitha Kraiprasit, BN*

** Department of Anesthesiology, Faculty of Medicine, Siriraj Hospital, Mahidol University*

The present study cross-validated self-report pain scales: Verbal Rating Scale (VRS), Facial Expression Scale (FACES), Color Analogue Scale (CAS) and Poker Chip Tool (PCT) in Thai children aged 5-12 years. The concordance with observational measure, Children's Hospital of Eastern Ontario Pain Scale (CHEOPS) was also tested. Among 100 students, test-retest reliability of all self-report measures was moderate to good (K = 0.501-0.712) and only FACES yielded acceptable face validity(IC > 0.5). Validation in 87 patients, all scales showed construct and concurrent validity. FACES was the most preferred scale. Agreement of self-report measures and CHEOPS was better in the age group 5-8 years (K = 0.417-0.826) than 9-12 years (K = 0.231-0.529). In conclusion, FACES is a valid, reliable and practical tool. Self-report measures are more in concordance with CHEOPS in the younger age group.

Keywords: *Cross-validation, Pain, Self-report measure, Concordance, CHEOPS*

J Med Assoc Thai 2005; 88(3): 412-8
Full text. e-Journal: *http://www.medassocthai.org/journal*

There has been an increase in research studying postoperative pain management in children. The use of valid and appropriate instruments to measure pain is very important to both clinical practice and research studies[1]. Pain measurement tools in children are age related, depending on the cognitive and language development of the child. As pain is a subjective experience, the self-report method is considered to be the gold standard for the assessment of pain and is suitable for school-age children (age $\geq$ 5 years)[2]. School-age children are capable of understanding verbal concepts and numbering, and are able to provide a more detailed rating of intensity and description of the quality and location of the pain[3].

Several self-report measures have been developed and validated for use with children. These are the categorical/ordinal Verbal Rating Scale (VRS)[4], Facial Expression Scale (FACES)[2], Color Analog Scale (CAS)[5] and Poker Chip Tool (PCT)[6].

Correspondence to : Suraseranivongse S, Department of Anesthesiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. E-mail: suwannee_18 @hotmail.com

Verbal Rating Scale is an ordinal 5-point pain scale, raging from no pain to most severe pain[4]. The Facial Expression Scale uses cartoons or actual photographs of facial expressions that reflect the universal reactions to pain. The number on pictures of each scale varies from 5 to 10[4]. The authors selected the Sheffield Children's Hospital Assessment tool which consists of 5 cartoon faces with adjectival descriptions of each face with a corresponding numerical scales (Fig. 1)[7]. The Colored Analog Scale is a ratio scale using gradations in color and area (wedge-shaped) along with a length score ranging from 0-10[5]. The Poker Chip Tool is a concrete ordinal rating tool consisting of 4 pieces representing pain, the number of chips corresponds with pain intensity[6]. This tool has been tested extensively and widely applied in clinical practice because of its simplicity and preference by children and nurses alike[8]. All pain scales have been validated in Western children. Culture may influence a child's perception, translation, cognition and preference[9], therefore, cross-validation should be performed before using each pain scale.

During the immediate postoperative period when children are not awake, nurses often rate pain

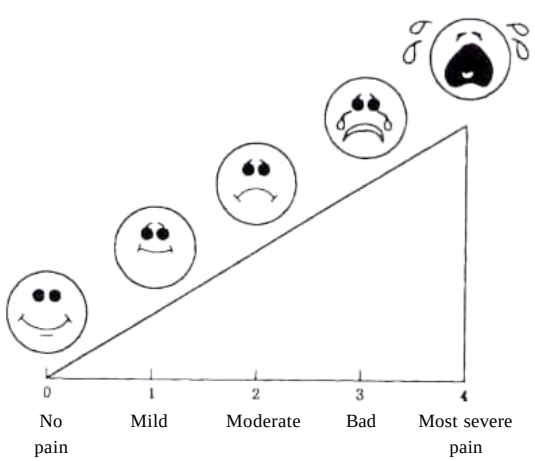

Fig. 1 Facial expression scale Sheffield Children's Hospital Assessment tool (with permission)

by observation. The relation between the self-reported and observer rating of pain are still controversial[10,11]. Pain tolerance is influenced by several factors including age, type of pain, type of surgery, postoperative stage and a child's coping style which are affected by culture and temperament.

The aims of the present study were 1) to cross-validate self report measures: VRS, FACES, CAS and PCT in school-age Thai children in terms of validity, reliability and practicality 2) to assess the concordance between the self-report scales and an observational scale (Children's Hospital of Eastern Ontario Pain scale).

Material and Method

After obtaining approval from the Research Ethics Committee and parental consent, school-aged children ages 5-12 years, ASA physical status I-II were enrolled. A Phase I study was conducted in healthy children in a primary school for testing of face validity and test-retest reliability. Phase II was conducted in patients undergoing general anesthesia and surgery at Siriraj Hospital, a tertiary hospital in Bangkok, Thailand to test the validity, practicality and concordance with an observational scale. Patients were excluded if they had postoperative ventilatory support, chronic pain or developmental delay.

Cross-validation was performed for translation, reliability testing, validity testing and practicality testing.

Translation

The six adjectives used to describe pain in the 4 measurement scales were "no pain, mild pain, moderate pain, bad, severe pain and most severe pain" were translated into Thai by 3 Anesthesiologists who were fluent in both Thai and English. Alterations were discussed for concensus.

Reliability testing

Reliability is a measure of consistency. Self-report tools depend on children's assessment of themselves, therefore only test-retest reliability could be tested. School-age children were asked to use the 4 tools to rank their pain experience of 2 scenarios of acute sharp pain namely needle injection, and the pinching of a finger from a door. The same procedure was then repeated 2 weeks after the first occasion.

Validity testing

Validity is a measure of accuracy, and was evaluated as follows:

Face validity

All pain scales were tested on the appearance to measure pain by children aged 5 to 12 years. Validity of each pain scale was scored globally (as 1 = relatively valid, 0 = not sure, -1 = relative irrelevant) on 2 occasions before and after using all scales.

Construct validity

This is an assessment of the meaning of the instrument in terms of its theoretical basis by comparison with external variables related to the construct. The authors compared the scores of all pain scales at no pain before surgery with those after surgery, before analgesia, as the postoperative pain scores were expected to be higher than those recorded preoperatively. Postoperative pain scores were also compared in patients with and without regional anaesthesia as the pain rating in patients who did not receive regional anaesthesia was expected to be higher than those who received regional anaesthesia.

Since CAS in the present study used a scale of 0 to 10 which is different from the 0-4 scale used in VRS, FACES, PCT. A CAS rating of ≥ 4 was used to differentiate pain corresponding to the VRS, FACES, PCT rating of $\geq$ moderate pain, which was considered as pain needing treatment.

Concurrent validity

The correlation of VRS, FACES, CAS and PCT were tested at the same point in time.

Practicality testing

Time spent in rating pain using each tool was recorded. Children were asked to select their favorite pain scale based on preference, appeal, utility and simplicity.

Relation between self-report scales and a behavioral observational scale

Pain behaviors were recorded by 5 research nurses using CHEOPS at the same time the self-reported scales were measured. Inter-rater reliability among nurses had been tested with high intraclass correlation (> 0.9).

Statistics

The sample size was calculated on the basis of a descriptive study with a variation of 10% and incidence of pain of 75%[12]. The formular $n = Z \alpha^2 pq/\delta^2$ as used, $\alpha = 0.5$, $p = 0.75$, $q = 1-p$, $\delta = 0.1$. The estimated sample size was 72. Demographic data were analyzed using descriptive statistics. Face validity was assessed for each scale using Item Correlation (IC), which is the summation of the total score of each scale divided by the number of evaluators. If the IC was ≥ 0.5, face validity would be accepted. Test-retest reliability was analyzed by intraclass correlation ICC) for CAS ($ICC = \delta^2$ subject/(δ^2 subject + δ^2 observer + δ^2 error)) and by Kappa statistics (K) for VRS, FACES and PCT. Kappa was analyzed based on grouping ratings of no pain and mild pain as the "low pain" and moderate to most severe pain as the "high pain" group. An ICC of ≥ 0.8 was considered acceptable. The value of K was interpreted as follows: < 0.2 = poor agreement, 0.21-0.4 = fair agreement, 0.41-0.6 = moderate agreement, 0.61-0.8 = good agreement, 0.81-1.0 = very good agreement.

Construct validity was analyzed using Wilcoxon Rank Sum test for non parametric data and Chi square test for categorical data. Concurrent validity was analyzed using Spearman correlation. The practicality of the scale such as time taken to rate pain score and the ranking of tools were analyzed with descriptive statistics. All analyses were performed with SPSS for window V.9 (SPSS, Chicago, IL, USA).

Results

In Phase I of the study 100 students, aged 6-12 years (mean 9.28 $\pm$ 1.49 years), 50% male, were enrolled (Table 1). Test-retest reliability was moderate to good (Table 2). Face validity assessed by students revealed that only FACES was acceptable both before and after using pain scales (IC > 0.5) (Table 3).

In Phase II of the study, of the 87 patients enrolled, age range 5-12 years (mean 8.27 $\pm$ 2.23 years)

Table 1. Demographic data, values are mean (SD) or number (proportion)

Characteristic	Value
Age: years	
- Students (n = 100)	9.28 (1.49)
- Patients (n = 87)	8.27 (2.23)
Age group	
- Students	
5-8 y	47 (54.0%)
9-12 y	40 (46.0%)
- Patients	
5-8 y	28 (28.0%)
9-12 y	72 (72.0%)
Sex (male)	
- Students (n = 100)	50 (50.0%)
- Patients (n = 87)	58 (66.7%)
Type of patients	
- Inpatients	58 (66.7%)
- Outpatients	29 (33.3%)
Type of surgery	
- Groin & perineum	37 (42.5%)
- Abdomen, kidney, ureter	23 (26.4%)
- Maxillofacial, head-neck	13 (14.9%)
- Extremity	6 (6.9%)
- Superficial	4 (4.6%)
- Ear,nose,throat	3 (3.1%)
- Eye	1 (1.1%)
Type of anesthesia	
- General only	32 (36.8%)
- General + regional	55 (63.2%)

Table 2. Test-retest reliability of the pain scales

Pain scales	Needle injection		Door-crushed injury	
	K	ICC	K	ICC
VRS	0.501	-	0.713	-
FACES	0.603	-	0.597	-
CAS	-	0.6320	-	0.7118
PCT	0.532	-	0.607	-

Table 3. Face validity

Pain scales	Face validity (Item correlation)	
	Before using pain scales	After using pain scales
VRS	0.40	0.47
FACES	0.71	0.64
CAS	0.32	0.21
PCT	0.26	0.23

who were recovering from various types of surgery (Table 1), 28 outpatients had missing ward data.

Construct validity clearly demonstrated a significant difference in pain scores (Table 4), and in the number of patients with moderate to most severe pain (Table 5) before and after surgery, before analgesia. Median pain scores on the ward were lower than median pain scores in the post anaesthetic care unit (PACU) (Table 4). Postoperative pain scores in the PACU of all pain scales in patients who received only general anaesthesia were higher than those who received supplemented regional anaesthesia (Table 6).

Concurrent validity was assessed in terms of correlation. All data before and after surgery, before analgesia, both in the PACU and on the ward were analyzed for association. The correlation of the 4 pain scales with each other was moderate to good ($r = 0.501$-0.821, $p < 0.0001$) before surgery and good ($r > 0.8$, $p < 0.0001$) after surgery both in the PACU and on the ward (Table 7). In terms of practicality, all pain scales were comparable for the duration of rating. FACES was the most preferred and VRS was the least preferred (Table 8).

Regarding the relationship between the self-reported and the observational pain scale (CHEOPS) for rating moderate to severe pain, agreement was moderate to good in the 5-8 year-old age group and fair to moderate in the older age group (Table 9).

Discussion

The present results indicated that all pain scales demonstrated moderate to good test-retest reliability and concurrent validity. All pain scales yielded better correlation of scores after surgery than before surgery. Construct validity of all scales was clearly demonstrated by a significantly higher pain score and proportion of patients in pain after surgery than before surgery. Of the 4 pain scales tested, FACES was found to be superior to the others regarding patients' preference and its acceptable face validity assessed by students. The present study also found

Table 4. Construct validity of pain scales before and after surgery in the PACU and on the ward. Values are median (IQR)

Pain scales	Pain scores before surgery (n = 87)	Pain scores after surgery		p-value
		PACU (n = 87)	Ward (n = 55)	
VRS	0 (0-0)	2 (0-3)	1 (0-3)	<0.0001
FACES	0 (0-0)	2 (1-2)	1 (0-2)	<0.0001
CAS	0 (0-0)	3 (0.5-5)	2 (0-4)	<0.0001
PCT	0 (0-0)	1 (1-2)	1 (0-2)	<0.0001

Wilcoxon Rank Sum test

Table 5. Construct validity of number of patients with moderate to most severe pain before and after surgery in the PACU and on the ward. Values are number (proportion)

Pain scales	Pain scores before surgery (n = 87)	Pain scores after surgery		p-value
		PACU (n = 87)	Ward (n = 55)	
VRS	0 (0%)	49 (56.3%)	27 (48.2%)	<0.0001
FACES	0 (0%)	46 (52.9%)	20 (36.4%)	<0.0001
CAS	0 (0%)	39 (45.3%)	15 (27.3%)	<0.0001
PCT	0 (0%)	43 (49.4%)	24 (42.9%)	<0.0001

Chi-square test

Table 6. Pain scores in the PACU after general anesthesia (GA) only or combined with regional anesthesia (GA+ RA). Values are median (IQR) or mean (SD)

Pain scales	Pain scores In PACU		p-value
	GA (n = 32)	GA + RA (n = 55)	
VRS	2 (1-3)	1 (0-2)	0.035*
FACES	2 (1-3)	1 (0-2)	0.042*
CAS	5.04 (3.41)	2.94 (3.22)	0.005#
PCT	2 (1-2.75)	1 (0-2)	0.051*

* Mann-Whitney U test, # unpaired t-test

Table 7. Concurrent validity (correlation) among pain scales before and after surgery in the PACU and on the ward. Spearman correlation ($p < 0.0001$)

Pain scales	Before surgery				PACU				Ward			
	VRS	FACES	CAS	PCT	VRS	FACES	CAS	PCT	VRS	FACES	CAS	PCT
VRS	1				1				1			
FACES	0.699	1			0.881	1			0.830	1		
CAS	0.821	0.567	1		0.891	0.852	1		0.872	0.867	1	
PCT	0.621	0.653	0.501	1	0.877	0.899	0.899	1	0.873	0.828	0.908	1

Table 8. Practicality of pain scales after surgery. Values are median (IQR) (range) or number (proportion)

Item	PACU				Ward			
	VRS	FACES	CAS	PCT	VRS	FACES	CAS	PCT
Time spent in rating	0 (0-0) [0-30]	0 (0-0) [0-15]	0 (0-0) [0-10]	0 (0-0) [0-20]	0 (0-0) [0-2]	0 (0-0) [0-20]	0 (0-0) [0-5]	0 (0-0) [0-15]
Pain scale of most preference	-	-	-	-	2 (2.6%)	45 (58.4%)	15 (19.5%)	15 (19.5%)

Table 9. Agreement between self-report measures (VRS, FACES, CAS, PCT) and an observational measure (CHEOPS) in rating moderate to most severe pain according to age group and postoperative stage. (Cut off point: VRS, FACES, PCT ≥ moderate, CAS ≥ 4, CHEOPS ≥ 8)

	Agreement with CHEOPS after surgery					
	PACU			Ward		
	All age	5-8y	9-12y	All age	5-8y	9-12y
VRS	0.360	0.571	0.231	0.380	0.417	0.349
FACES	0.459	0.657	0.335	0.656	0.826	0.497
CAS	0.512	0.620	0.410	0.700	0.811	0.575
PCT	0.433	0.582	0.356	0.456	0.589	0.349

Kappa statistics

more concordance between the self-report scales and CHEOPS in the younger age group (5-8 year) than the older age group (9-12 year).

Self-report measures do not address or reflect nociception, but rather the experience of pain which is complex and difficult to measure as a unidimensional construct. Pain is a multi-dimensional experience which includes sensory intensity, quality, emotional effect or affective magnitude, location of daily living, time course, interference with function at work and activities of daily living, sleep and social interaction[2]. Self-report of pain is a behaviour which can be influenced by a range of cognitive, cultural, emotional and motivational states. In adults, a ratio scale is assumed to provide the optimal measure of clinically meaningful changes in pain levels. But in paediatrics, the virtues and precision of ratio scale may be less important than the scale's simplicity and appeal to a child.

The Sheffield Children's Hospital Assessment Tool which relies on assessment of facial expression was a valid, reliable and practical tool based on the presented data. There may be several reasons for this. First, this measure consists of cartoon faces with graded facial responses to pain combined with a graded height triangle relating to the severity of pain, an adjectival description along with a numerical scale. All techniques in this measure might increase the ability of children to discern the face that reflected their feelings. Second, there is considerable general support for cross-cultural consistency in facial expressions of pain[13,14]. Third, a facial expression scale truly reflects the universal reaction to pain. It is simple to use, readily understood and likely to be developmentally appropriate for young children[15]. Fourth, most children prefer cartoons regardless of age, gender or race.

The Poker Chip Tool has been extensive tested and widely applied in western children. It has been recommended by the Task Force on Acute Pain of the IASP for use in children as young as 4 years of age[16]. However, the present findings were contrary. Many of the students did not accept the appearance of this scale as a pain assessment tool. They might find it difficult to reflect the intensity of their feeling of pain using the pieces of hurt in PCT. In addition, the Color Analogue Scale which possesses a ratio scale quality was assumed to be easier than a linear Visual Analogue Scale[2]. However, the appearance of the CAS not understandable for use in measurement pain in the presented patients. Relating intensity of pain feeling to varying degrees of color, width and length may be difficult for Thai children.

Verbal Rating Scales (adjectival descriptor) is suggested as being appropriate for children about 12 years old or older[2]. The present results also supported this suggestion because the face validity among students and preference assessment among patients 5-12 years old were poor. This might be attributed to a language skill barrier, i.e. the scale was not simple enough to explain their feeling of pain.

Agreement between self-reported and observational pain scales of postoperative pain in the present study showed more concordance in the younger age group (5-8 years) than the older age group (9-12 years). The present finding supported a previous study which reported that the correlation between self-report of needle pain and observable behaviors was significantly reduced with increasing

age[2]. Therefore, asking a patient's feeling is necessary for postoperative pain assessment in older children even though they may look calm and quiet.

In conclusion, on the basis of the present data from school-age Thai children, the Facial Expression Pain Scale, "Sheffield Children's Hospital Assessment tool", was the most valid, reliable and practical measure for postoperative pain assessment. The relationship between self-reported and observational pain scales were in more concordance in the children from the younger age group (5-8 years).

Acknowledgements

The authors wish to thank Dr. JM Goddard and Blackwell Publishing Ltd, Professor Patricia A McGrath and Professor Nancy Hester who gave permission to cross-validate Sheffield Children's Hospital Assessment tool**,** Color Analogue Scale and Poker Chip Tool respectively. Financial support was provided by the Siriraj Research Development Fund and Senior Research Scholar, the Thailand Research Fund.

References

1. Finley GA, McGrath PJ. Introduction: The roles of measurement in pain management and research. In: Finley GA, McGrath PJ, eds. Measurement of pain in infants and children. Seattle: IASP Press, 1998; 10: 1-4.
2. Champion GD, Goodenough B, von Baeyer CL, Thomas W. Measurement of pain by self-report. In: Finley GA, McGrath PJ, eds. Measurement of pain in infants and children. Seattle: IASP Press, 1998; 10: 123-60.
3. Franck LS, Greenburg CS, Stevens B. Pain assessment in infants and children. Pediatr Clin North Am 2000; 47: 487-512.
4. Savedra MC, Tesler MD. Assessing children's and adolescents' pain. Pediatrics 1989; 16: 24-9.
5. McGrath PJ, 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: 435-53.
6. Hester NK. The preoperational child's reaction to immunization. Nurs Res 1979; 28: 250-5.
7. Goddard JM, Pickup SE. Postoperative pain in children: combining audit and a clinical nurse specialist to improve management. Anaesthesia 1996; 51: 588-90.
8. Hester NO, Foster R, Kristensen K. Measurement of pain in children: generalizability and validity of the Pain Ladder and the Poker Chip Tool. In: Tyler DC, Krane EJ, eds. Pediatric pain. Advances in pain research and therapy. New York: Raven Press, 1990; 15: 79-84.
9. Chen E, Joseph MH, Zelter K. Behavioral and cognitive interventions in the treatment of pain in children. Pediatr Clin North Am 2000; 47: 513-25.
10. Beyer JE, McGrath PJ, Berde CB. Disconcordance between self –report and behavioural pain measures in children aged 3-7 years after surgery. J Pain Symptom Manage 1990; 5: 350-6.
11. Tyler DC, Tu A, Douthit J, Chapman CR. Toward validation of pain measurement tools for children: a pilot study. Pain 1993; 53: 301-9.
12. Mather L, Mackie J. The incidence of postoperative pain in children. Pain 1983. 15: 271-82.
13. LeResche L, Dworkin SF, Wilson L, Ehrlich KJ. Effeect of temporomandibular disorder pain duration on facial expression and verbal reports of pain. Pain 1992; 51: 289-95.
14. Craig JD. The facial display of pain. In: Finley GA, McGrath PJ, eds. Measurement of pain in infants and children. Seattle: IASP Press, 1998; 10: 103-21.
15. Goodenough B, Addicoat L, Champion GD, McInerney M, Young B, Juniper K, et al. Pain in 4- to 6-year-old children receiving intramuscular injections: a comparison of the Faces Pain Scale with other self-report and behavioral measures. Clin J Pain 1997; 13 60-73.
16. Ready LB, Edwards WJ. Management of acute pain. Seattle: IASP Publication, 1992.
17. Goodenough B, Champion GD, Labreaux L, Tabah L, Kempel L. Needle pain severity in children: does the relationship between self-report and observed behavior vary as a function of age? Aust J Psychol 1998; 50 1-9.

การตรวจสอบความเที่ยงตรงของเครื่องมือวัดระดับความปวดชนิด self-report หลังผ่าตัดในเด็กไทยวัยเรียน

สุวรรณี สุรเศรณีวงศ์, ธัญนิตย์ มันตภาณีวัฒน์, จิตประภา มานนท์, เพชรี เจนจบ, สุพพัต เพชรรัตน์, ขนิษฐา ไกรประสิทธิ์

การศึกษานี้มีวัตถุประสงค์เพื่อ 1) ตรวจสอบความเที่ยงตรงของเครื่องมือวัดระดับความปวดชนิดที่ผู้ป่วยประเมินความรู้สึกของตนเอง (self-report) 4 ชนิดได้แก่ Verbal Rating Scale (VRS), Facial Expression Scale "Sheffield Children's Hospital Assessment tool" (FACES), Color Analog Scale (CAS) และ Poker Chip Tool (PCT) ในเด็กไทยอายุ 5-12 ปีหลังผ่าตัด 2) ศึกษาความสัมพันธ์ระหว่างเครื่องมือวัดระดับความปวดชนิด self-report และชนิดให้ผู้อื่นประเมินโดยสังเกตอาการ (observational measure) โดยใช้ Children Hospital of Eastern Ontario pain Scale (CHEOPS) ผลการศึกษาในเด็กนักเรียน 100 คน พบว่า Test-retest reliability ของเครื่องมือ วัดระดับความปวดทั้ง 4 ชนิดอยู่ในระดับปานกลางถึงดี (K = 0.501-0.712) และมีเพียง FACES เท่านั้น ที่มี face validity อยู่ในระดับที่ยอมรับได้ (IC > 0.5) สำหรับการศึกษาในผู้ป่วยจำนวน 87 ราย พบว่าเครื่องมือทั้ง 4 ชนิด มี construct และ concurrent validity โดยที่ผู้ป่วยชอบใช้ FACES มากที่สุด ส่วนความสัมพันธ์ระหว่างคะแนนความปวดที่ผู้ป่วยประเมินตนเองด้วยเครื่องมือทั้ง 4 ชนิด กับคะแนนที่พยาบาลประเมินโดยใช้ CHEOPS พบว่าเป็นไปในแนวทางเดียวกันในกลุ่มอายุ 5-8 ปี (K = 0.417-0.826) มากกว่ากลุ่มอายุ 9-12 ปี (K = 0.231-0.529) สรุป FACES เป็นเครื่องมือที่มีความเที่ยงตรง และเหมาะสำหรับใช้ในเด็กไทย การประเมินระดับความปวดโดยตัวเด็กเองจะใกล้เคียงการประเมินจากพยาบาลในกลุ่มอายุ 5-8 ปี มากกว่า 9-12 ปี