Journal of Pain Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

A Systematic Review of the Measurement Properties of Face, Legs, Activity, Cry and Consolability Scale for Pediatric Pain Assessment

Tuochao Peng, Shuangquan Qu, Zhen Du, Zheng Chen, Ting Xiao, Ru Chen

Department of Anesthesia and Surgery, Hunan Children's Hospital, Changsha, 410007, People's Republic of China

Correspondence: Tuochao Peng, Department of Anesthesia and Surgery, Hunan Children's Hospital, Changsha, 410007, People's Republic of China, Tel +86 13755074295, Email 258014811@qq.com

Purpose: We performed this systematic review to summarize the psychometric properties of Face, Legs, Activity, Cry and Consolability (*FLACC*) scale in pediatric patients in different settings.

Methods: Two investigators independently searched PubMed, EMBASE, OVID and China National Knowledge Infrastructure (CNKI) for eligible studies through July 2021. We assessed the psychometric properties using the modified critical appraisal tool (CAT). Finally, we systematically reviewed the results of the included studies.

Results: A total of 15 studies were eventually included. The overall quality of each eligible study was low to moderate. The *FLACC* scale has been available in different versions and in different settings. Although eligible studies have demonstrated significant clinical benefit in assessing postoperative pain in pediatric patients aged 0 to 10 years from post-anesthetic care unit (PACU), pediatric intensive care unit (PICU) and inpatient unit, and in assessing procedural pain in pediatric patients aged 0.5 to 7 years from emergency unit, immunization center and PICU, mostly without test-retest analysis.

Conclusion: Although the absence of a gold standard of pain assessment, the currently available data support the usefulness of the *FLACC* from the perspective of criterion validity. Therefore, the *FLACC* scale can be considered for measuring observational pain in infants and children. However, further studies are still needed to provide more robust evidence.

Keywords: children, *FLACC* scale, infants, pain assessment, systematic review

Introduction

In the pediatric unit, pediatric intensive care unit (PICU), or emergency unit, infants and children undergoing medical examinations and treatments often experience all kinds of pain, especially procedural pain, including pain from blood collection, venipuncture, injections and even surgery.[1] However, infants and children often lack the verbal and cognitive skills to complete self-reports of physical discomfort or pain intensity, leading to inadequate pain assessment and subsequent treatment in infants and children.[2] Therefore, it is critical to have effective and reliable instruments for timely and accurate assessment of pain in infants and children.

Currently, more than 40 multidimensional observational scales for assessing pain intensity have been developed, some of which have been adapted and applied to different populations and clinical settings.[3–6] Generally, these scales are combinations of specific distressing behaviors, including crying, facial expression, body movement, and more. The Face, Legs, Activity, Cry and Consolability (*FLACC*) scale was published in 1997 to assess postoperative pain in children and is currently one of the most commonly used scales.[7] The *FLACC* scale contains 5 behaviors, including face, legs, activity, consolability, and cry, and each behavior is scored from 0 to 10. Compared with other scales assessing pain in infants and children, the *FLACC* scale is more convenient and practical in identifying and recording pain.[8–10]

Although the original study that developed the *FLACC* scale analyzed psychometric properties, the study was at risk of bias because raters were not blinded to the use of analgesics.[7] To date, numerous studies have attempted to evaluate

Received: 10 November 2022
Accepted: 21 March 2023
Published: 8 April 2023

the reliability and validity of this scale. Specifically, some focus on the application of different groups of people such as neonates, preschool children, and children with cognitive impairment, and some focus on the application in different settings, such as acute pain, postoperative pain, and procedural pain.[11–14] Studies published before 2015 have been included in 3 independent systematic reviews.[8–10] Based on the currently available evidence, all 3 systematic reviews temporarily recommended the use of the *FLACC* scale for the assessment of pain in children. However, the conclusions of these previous 3 systematic reviews should be interpreted cautiously because these systematic reviews did not differentiate the clinical setting in which the *FLACC* scale was applied or define age ranges for pediatrics. In addition, several new methodological studies have been published since the publication of the latest systematic review.

Therefore, to provide more accurate and robust evidence for the clinical use of the *FLACC* scale by addressing the limitations of all previous systematic reviews, the present systematic review was performed to include currently available studies to further comprehensively summarize the reliability, validity, responsiveness, and feasibility of the *FLACC* scale for assessing pain in infants and children without cognitive impairment for a specific age range (0–12 years) in different settings through including currently available studies.

Materials and Methods

This systematic review was designed according to the Cochrane Collaboration (CC). Findings were reported according to the preferred items for systematic review and meta-analysis (PRISMA) checklist.[15] Ethic approval was not required for this study as it was a systematic review of published studies.

Identification of Citations

Two investigators searched four electronic databases, including PubMed, Embase, OVID and China National Knowledge Infrastructure (CNKI) for relevant studies through July 2021. Search results are updated weekly to find any potential studies. The search strategy was designed using "Face, Legs, Activity, Cry and Consolability", "child", "neonate" and "infant". The search strategy of PubMed is shown in Table S1. Meanwhile, we manually checked the references of all included studies and topic-related reviews. Finally, we managed all records with EndNote X7.

Selection Criteria

We included only original methodology articles that primarily confirmed the psychometric evidence for both original and culturally adapted versions of the *FLACC* scale, such as its reliability, validity, responsiveness, or feasibility in infants and children aged 0–12 years and had no cognitive impairment. Studies that included children over the age of 12 years were not included. Meanwhile, we excluded the clinical controlled trials, animal experiments, reviews, guidelines, commentaries, and unpublished manuscripts. We also excluded other descriptive articles when the *FLACC* scale was the reference scale for examining the convergent validity. Only studies published in English or Chinese were considered eligible.

Data Extraction

Two investigators independently extracted the first author's name, publication year, country, sample size, pain type and setting, the psychometric properties including reliability, validity, responsiveness and feasibility, and the index related to pain diagnosis. Any conflicts regarding data extraction were resolved through discussion with a third investigator.

Assessment of Methodological Quality

We assessed psychometric evidence for the *FLACC* scale using 8 items from the critical appraisal tool (CAT) developed by Brink et al in 2012.[16] The evaluation was the part of the scope of this systematic review, not the exclusion criterion of studies and the evaluation of the validity of the results of each included studies. Given that there is no gold standard for observational pain assessment, therefore we excluded all items which were developed to focus on this aspect from our methodological quality assessment. In addition, using the unified procedure for data collection in different studies is impossible. Therefore, a total of 5 items, including item 3, 6, 7, 9, and 10, were excluded. Eventually, we used 8 items to assess the psychometric properties of the *FLACC* scale, including item 1 (if human subjects were used, did the authors

give a detailed description of the sample of subjects used to perform the test?), item 2 (did the authors clarify the qualification, or competence of the raters who performed the test?), item 4 (if inter-rater reliability was tested, were raters blinded to the findings of other raters?), item 5 (if inter-rater reliability was tested, were raters blinded to their prior findings of the test under evaluation?), item 8 (was the stability of the variable being measured taken into account when determining the suitability of the time interval between repeated measures?), item 10 (was the execution of the test described in sufficient detail to permit replication of the test?), item 12 (were withdrawals from the study explained?), and item 13 (were the statistical methods appropriate for the study?).

Assessment of Validity and Reliability Data

We performed a systematic review to assess the essential information of the *FLACC* scale from the following 2 aspects: psychometric assessment and target population. For psychometric properties assessment, we systematically evaluated the results of each eligible study from the following four aspects, reliability, validity, responsiveness, and feasibility. Specifically, Cronbach's alpha (α) coefficient, test-retest reliability coefficient, intra-class correlation coefficient (*ICC*), and *Kappa* coefficient higher or equal to 0.75 were considered good.[17,18] The content validity index of the overall scale (S-CVI) > 0.90 and an item's content validity index (I-CVI) ≥ 0.78 indicated that content validity in this study was satisfactory in the instrument.[19] The convergent validity was low, moderate, and large if the relationship coefficient (r) was <0.4, 0.4 to 0.7, and >0.7, respectively.[20] In addition, separate systematic reviews were designed according to different kinds of pain, including postoperative and procedural pain. In addition, we graphically displayed the methodological quality of each included study using Microsoft Excel. According to the actual information each eligible reported, the quality of each item was rated as "yes", "not applicable", or "no". By referencing the criteria of the Cochrane risk of bias assessment, the overall quality of a study was rated as "high" if it was rated "yes" in all items; as "low" if it was rated "no" in at least one item; as "moderate" if it was rated "not applicable" in at least one item but was not rated "low" in anyone.

Results

Result of Identification

By searching 4 electronic databases, we captured 1685 records. After the first selection and further full-text selection, 15 records were included in this study.[6,7,11–14,21–29] The flow diagram of article retrieval and selection is displayed in Figure 1.

Characteristics of Included Studies

Five studies were performed in America, 3 in China, 2 in Australia, and the remaining in other countries such as Japan, Sweden, Greece, France, and Thailand. All studies were published between 1997 and 2021. The *FLACC* scale was translated into different languages such as Chinese, Thai, Japanese, Swedish and Greek. The sample size of the 15 included studies ranged from 20 to 170. Four of the 15 included studies used the *FLACC* scale to assess procedural pain in immunization center, emergency unit and PICU.[6,25,27,29] Only one study assessed acute pain in the emergency unit.[26] Ten other studies assessed postoperative pain in different setting, such as post-anesthetic care unit (PACU), PICU, cardiac intensive care unit (CICU), inpatient unit, and burns/surgical/trauma unit.[7,11–14,21–24,28] All 15 included studies demonstrated the reliability and 12 studies demonstrated the validity. Nine studies provided evidence of responsiveness, and only 3 studies examined the feasibility of the *FLACC* scale. The basic characteristics of the 15 studies are displayed in Table 1.

Assessment of Psychometric Properties

The overall methodological quality of the 15 included studies were rated as moderate because most of all included studies had insufficient test-retest reliability analyses, and most did not assess intra-rater reliability. More details of the assessment results are presented in Figure 2.

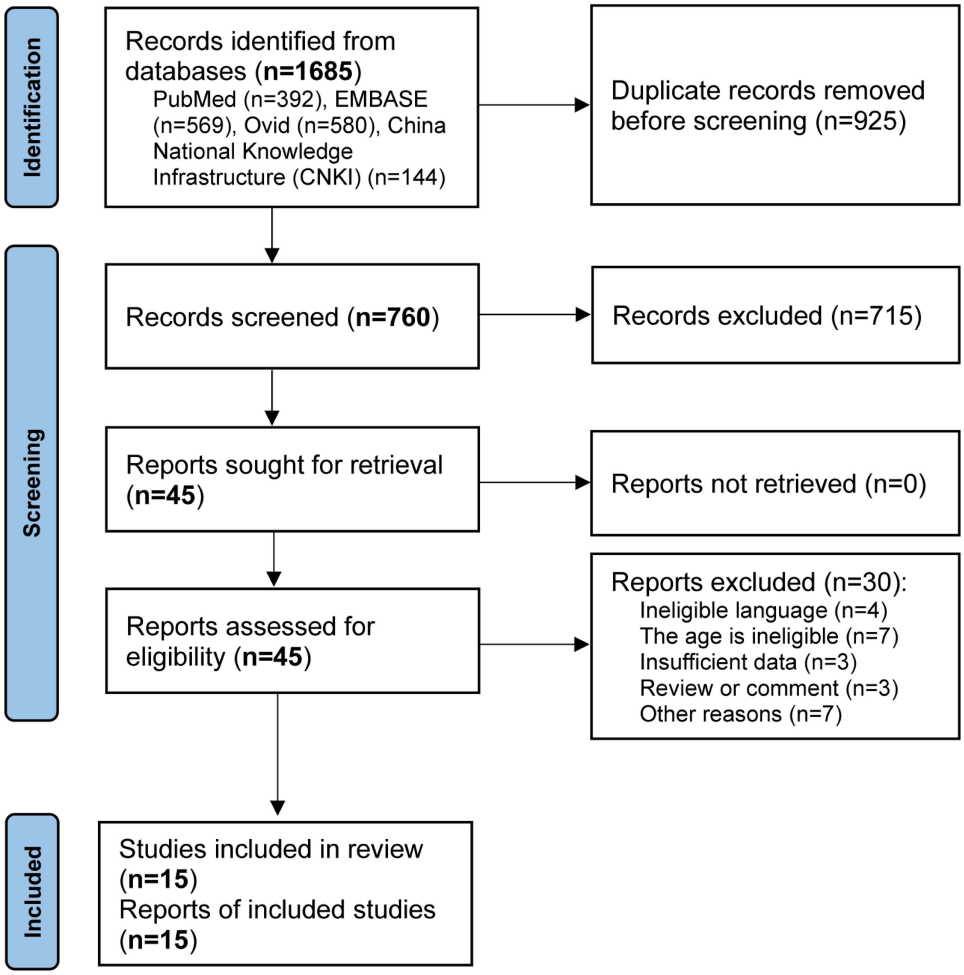

Figure 1 Flow diagram of retrieval and selection of eligible studies.

Systematic Review of Psychometric Properties of the *FLACC* Scale

Reliability

Assessment of Postoperative Pain in Different Settings

Ten studies examined reliability in infants and children with postoperative pain, including interrater reliability, intra-rater reliability, internal consistency, or test-retest reliability. Only 2 studies reported test-retest reliability.[14,24]

Merkel et al in 1997 developed the original scale and they assessed reliability in 89 postoperative infants and children aged 2 months to 7 years.[7] Psychometric test showed good inter-rater reliability of 0.97. Suraseranivongse et al in 2001 used the videotapes of 167 children aged 1 to 5 years from PACU to perform the psychometric test, indicating good inter-rater reliability of 0.949 and good intra-rater reliability among four observers ranged from 0.950 to 0.991.[21] In 2003, Manworren et al enrolled 147 infants and children aged 0 to 3 years from 5 units including PICU, PACU, surgical/trauma unit, hematology/oncology unit, and infant unit, indicating a moderate interrater reliability of 0.61 among 19 nurse volunteers.[22] Willis et al found 100% agreement among 6 pairs of observations, the limited quality of the study could not support good inter-rater reliability.[23] Bringuier et al in 2009 completed the psychometric evaluation by video records from 148 infants and children aged 1 to 7 years in inpatient unit, suggesting good interrater reliability (*ICC*>0.86) and internal consistency (Cronbach'α = 0.93).[11] Meanwhile, Johansson et al in 2009 only evaluated the inter-rater reliability by surveying 40 intubated and ventilated infants and children aged 0 to 10 years from PICU, revealing moderate *kappa* coefficient of 0.63.[12] In 2012, Bai et al tested the inter-rater reliability, indicating the ICC value was 0.84 by surveying two investigators among 4 critically ill children aged 0 to 10 years in CICU.[13] Jia et al enrolling 20 infants and children

Table 1 Basic Characteristics of Included Studies

Reference	Country	No.	Age	Circumstances	Scales	Diagnosis of Pain			Psychometric Properties			
						Cut-off	Sen	Spe	Reliability	Validity	Responsiveness	Feasibility
Merkel 1997[7]	America	89	2m-7y	Postoperative pain, PACU after surgery	Original FLACC	NA	NA	NA	Interrater reliability	Convergent validity (OPS)	Difference of pain scores before and after analgesia	NA
Suraseranivongse 2001[21]	Thailand	167	1–5y	Pain after awakening from general anesthesia until 24 h after surgery	Original FLACC translated into Thai	2	88%	86%	Interrater reliability, intra-rater reliability	Content validity, convergent validity (OPS, TPPPS, CHEOPS)	Difference of pain scores between immediately in the PACU and several hours later on the ward) and between age group (<3 and >3y)	Use in clinical situations, ease of use, ability of the scales to help assess pain, and general satisfaction with the scales
Manworren 2003[22]	America	147	0–3y	Acute and postoperative pain, 5 units including PICU, PACU, surgical /trauma unit, hematology/ oncology unit, and infant unit	Original FLACC	NA	NA	NA	Interrater reliability	NA	Difference of pain scores before and after analgesia	NA
Willis 2003[23]	America	30	3–7y	Postoperative pain, inpatient unit	Original FLACC	NA	NA	NA	Interrater reliability	Convergent validity (FACES)	NA	NA
Bringuier 2009[11]	France	148	1–7y	Postoperative pain, inpatient unit	Original FLACC	3	77%	96%	Interrater reliability, internal consistency	Face validity, construct validity, convergent validity (CHEOPS, CHIPPS, OPS), discriminant validity (VAS-anxiety)	Difference of pain scores at four times	NA
Johansson 2009[12]	Sweden	40	0–10y	Postoperative pain in intubated and ventilated children, PICU	Original FLACC translated into Swedish	NA	NA	NA	Interrater reliability	Convergent validity (COMFORT-B, VASobs)	NA	NA

(*Continued*)

https://doi.org/10.2147/JPR.S397064 DovePress

Table 1 (Continued).

Reference	Country	No.	Age	Circumstances	Scales	Diagnosis of Pain			Psychometric Properties			
						Cut-off	Sen	Spe	Reliability	Validity	Responsiveness	Feasibility
Bai 2012[13]	China	170	0–7y	postoperative pain in critically ill children, CICU	Original FLACC translated into Chinese	2	98%	88%	Interrater reliability (test before study)	Convergent validity (COMFORT-B, VASobs)	NA	NA
Jia 2012[14]	China	20	0–5y	Pain after children with burns, surgical/trauma unit	Original FLACC translated into Chinese	NA	NA	NA	Interrater reliability, internal consistency, test-retest reliability (a second assessment after 2 months)	Convergent validity (COMFOR, POCIS)	Difference of pain scores in the different process dealing with burn wound	NA
Liu 2012[24]	China	100	0–3y	Postoperative pain in children with cleft lip and palate, inpatient unit	Original LACC translated into Chinese	NA	NA	NA	Interrater reliability, internal consistency, test-retest reliability	Content validity, construct validity	Difference of pain scores before and after analgesia	NA
Gomez 2013[25]	Australia	29	12–18m	Procedural pain, immunization center	Original FLACC	NA	NA	NA	Interrater reliability, interrater reliability	NA	NA	NA
Kochman 2017[26]	America	101	6m-5y	Acute pain, emergency unit	Original FLACC	NA	NA	NA	Interrater reliability	NA	Difference of pain scores at different times	NA
Crellin 2018[27]	Australia	100	6–42m	Procedural pain, emergency unit	Original FLACC	2	94.90%	72.50%	Interrater reliability, interrater reliability	Convergent validity (VASobs- pain and VASobs- distress)	Differences-in-difference (DID) between FLACC score for painful vs non-painful procedure	Assessment of how easy the scale is to use and how well it performs
Matsuishi 2018[28]	Japan	24	38m	Postoperative pain, PICU	Original FLACC translated into Japanese	NA	NA	NA	Interrater reliability	Convergent validity (VASobs)	NA	NA

Tamvaki 2020[6]	Greece	30	4.1y	Procedural pain, PICU	Original FLACC translated into Greek	NA	NA	NA	Interrater reliability, internal consistency	Convergent validity (Comfort-B, BPS, VASobs)	NA	NA
Tsze 2021[29]	America	20	1–7y	Procedural pain, pediatric emergency unit	Original FLACC	NA	NA	NA	Interrater reliability, internal consistency	Convergent validity (OSBD-R, CHEOPS)	Difference of pain scores during baseline and administration phases	NA

Abbreviations: No, sample size; NA, no report; PICU, pediatric intensive care unit; PACU, post-anesthetic care unit; CICU, cardiac intensive care unit; OPS, the Objective Pain Scale; TPPPS, Toddler Preschool Postoperative Pain Scale; CHEOPS, Children's Hospital of Eastern Ontario Pain Scale; C-BIIP; OSBD-R, the Observational Scale of Behavioral Distress–Revised; Comfort-B, Comfort Behavior scale; BPS, the Behavioral Pain Scale; VASobs, Visual Analogue Scale observer scale; POCIS, Pain Observation Scale for Young Children; FACES, Wong-Baker FACES Pain Rating Scale; Sen, sensitivity; Spe, Specificity.

https://doi.org/10.2147/JPR.S397064 DovePress

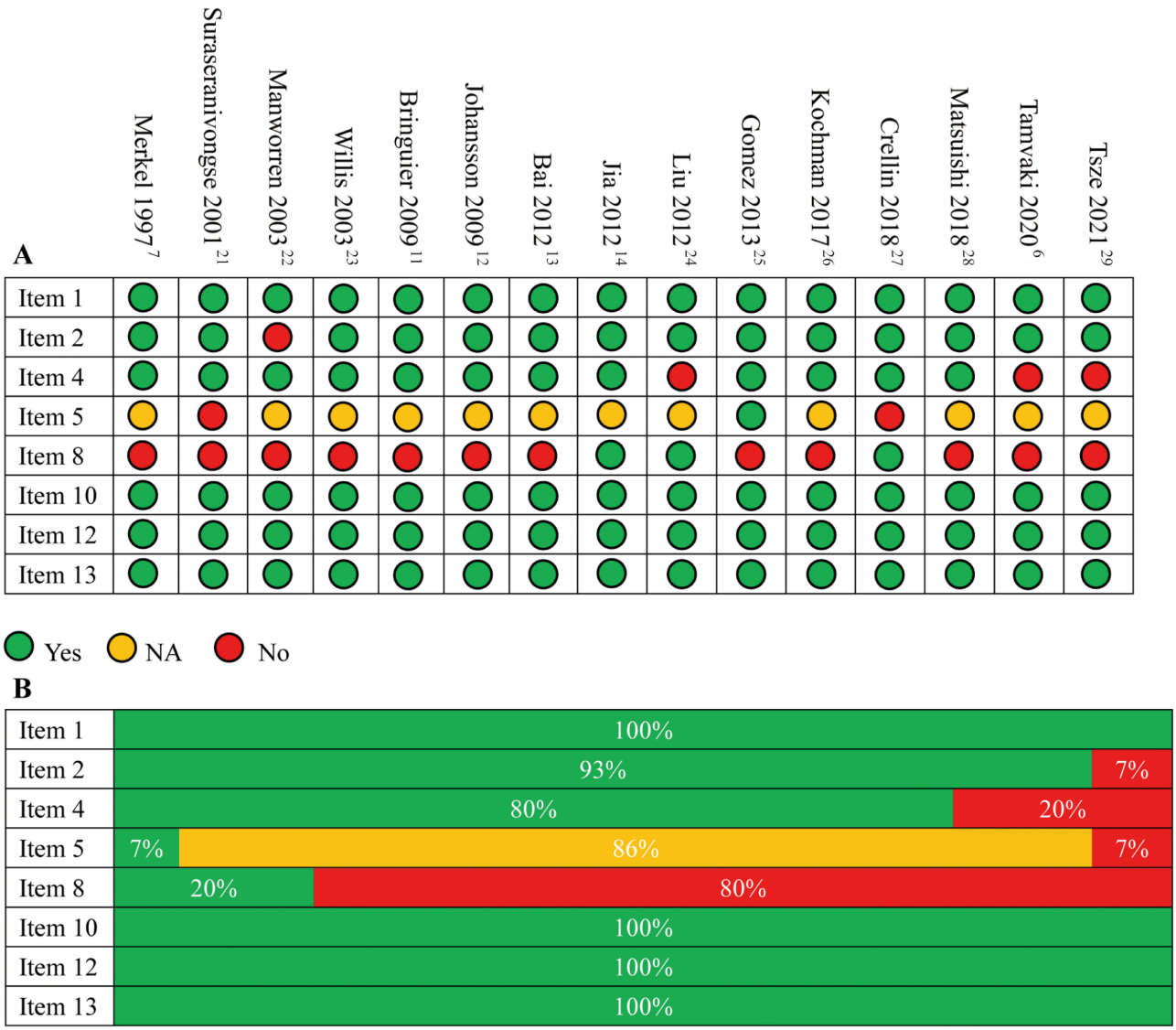

Figure 2 Assessment of reliability and validity of all the included studies (**A**) assessment graph and (**B**) assessment summary. In this figure, "yes (green)" indicates a study reported sufficient information required by an item, "no (red)" indicates a study did not report information required by an item, and "NA (yellow)" indicates that a study did not insufficient information to judge whether it meet the requirement of an item.

aged 0 to 5 years in burns/surgical/trauma unit and Liu et al enrolling 100 infants and children with cleft lip and palate aged 0 to 3 years in inpatient unit both finished the psychometric test including interrater reliability, internal consistency, and test-retest reliability.[14,24] The internal consistency in the study by Jia et al and Liu et al was not high, with a Cronbach'α of 0.790 and Cronbach'α of 0.745, respectively, and the test-retest reliability was $r = 0.645$ and $r = 0.946$, respectively. Liu et al only reported the interrater reliability was statistically significant and did not offer specific data. Jia et al showed the interrater reliability was 0.82. In 2018, Matsuishi et al involving 24 children in PICU only displayed excellent interrater reliability (*Kappa* value = 0.95).[28]

Assessment of Procedural Pain in Different Settings and Acute Pain

Only one study assessed acute pain in children, while 4 included studies assessed procedural pain in different settings.[6,25–27,29]

Gomez et al recruited 29 infants aged 12 to 18 months in immunization center and reported near-perfect inter-rater reliability ranging from 0.81 to 0.97 and intra-rater reliability ranging from 0.40 to 0.95 at different phases. Crellin et al

in 2018 surveyed 100 infants and children aged 6 to 42 months in the emergency unit by reviewing videos. They found high inter-rater reliability at different phases ranging from 0.79 to 0.94 and equally high intra-rater reliability (*ICC* = 0.87). Tamvaki et al in 2020 surveyed 30 children in PICU and Tsze et al in 2021 surveyed 20 children aged 1 to 7 years in pediatric emergency unit. Both offered high inter-rater reliability (*ICC* > 0.96 and *ICC* = 0.98, respectively) and high internal consistency (Cronbach'α > 0.88 and Cronbach'α = 0.97, respectively.

In 2017, Kochman et al assessed acute pain in 101 infants and children aged 6 months to 5 years in the emergency unit. They revealed strong inter-rater reliability (*Kappa* value ranged from 0.85 to 0.96).

Validity

Assessment of Postoperative Pain in Different Settings

Total 9 included studies reported convergent validity, content validity, face validity, and discriminant validity in infants and children with postoperative pain.

Merkel et al found a high correlation between the *FLACC* scale and OPS (r = 0.800).[7] In contrast, Suraseranivongse et al found moderate correlation between the *FLACC* scale and OPS (r = 0.721), the *FLACC* scale and CHEOPS (r = 0.696), and the *FLACC* scale and TPPPS (r = 0.621), but good content validity assessed by the experts.[21] Willis et al also provided evidence of moderate convergent validity because the correlation between *FLACC* and FACES was r = 0.584.[23] Bringuier et al reported equivocal evidence for the validity of the *FLACC* scale. *FLACC* was highly correlated with 3 scales, CHEOPS, CHIPPS and OPS (r = 0.88–0.94) and FASS (r = 0.71–0.78). However, the *FLACC* score was also moderately correlated with VAS-anxiety, suggesting a limited ability to distinguish pain from anxiety.[11] Johansson et al found a moderate correlation between *FLACC* and VASobs (r = 0.50), but a good correlation between *FLACC* and COMFORT-B (r = 0.76).[12] Different from Johansson et al, Bai et al found a correlation between *FLACC* and the two scales including VASobs and COMFORT-B (with a r = 0.86 and r = 0.51, respectively).[13] Jia et al found an excellent correlation between *FLACC* and the two scales including COMFORT and POCIS (with a r = 0.958 and r = 0.872, respectively).[14] Similarly, Matsuishi et al found a high correlation between FLACC and VASobs (r = 0.96).[28] Liu et al revealed a good content validity through expert assessment, with the overall content validity index of 1.000 and the average of content validity indices of 1.000 for individual items.[24]

Assessment of Procedural Pain in Different Settings

Only 3 included studies reported convergent validity in infants and children with procedural pain.[6,27,29]

Crellin et al found that the correlation between the *FLACC* scale and VASobs pain and VASbs distress was higher for distress (r = 0.89) and pain (r = 0.74), respectively.[27] Tamvaki et al found that the correlation between *FLACC* scale and the 3 scales including Comfort-B, BPS and VASobs was high (r > 0.71).[6] Similar results were obtained by Tsze et al that the correlation between the *FLACC* scale and 2 scales including OSBD-R, CHEOPS was high (with a r = 0.88 and r = 0.96, respectively).[29] Overall, these 3 studies revealed good convergent validity for the use of *FLACC* in infants and children with procedural pain.

Responsiveness

A total of 9 studies of all included studies reported the assessment of responsiveness. A total of 6 studies assessed the responsiveness in infants and children with postoperative pain, 2 studies assessed procedural pain in infants and children, and 1 study assessed acute pain in infants and children.[7,11,14,21,22,24,26,27,29] The results of 9 studies provided proper evidence for the responsiveness of the *FLACC* scale, which were summarized in Table 1.

Feasibility

Two studies of all included studies tested the feasibility of the *FLACC* scale. Suraseranivongse et al evaluated the duration of rating, ease of use, ability of pain evaluation, and satisfaction with the *FLACC* scale. They found that *FLACC* scale tool less time to assess pain and was feasible for clinicians.[21] Crellin et al found the *FLACC* scale was easy to understand and use, but not very useful to assess assessing pain using a 0–2 point score out of 5 categories.[27]

Discussion

The *FLACC* scale is commonly used in infants and children for the measurement of three types of pain, including procedural pain, postoperative pain and acute pain. The original *FLACC* scale has been culturally adapted into different versions for use in different clinical settings. However, the reliability and validity of the *FLACC* scale have not been well confirmed in infants and children in different clinical settings. The current systematic review to further summarize the reliability and validity of the *FLACC* scale by only including studies primarily assessing the psychometric properties of the *FLACC* scale and excluding clinical studies, reviews and the descriptive studies which *FLACC* scale was utilized to assess the convergent validity of other pain scales. In the current systematic review, we captured 15 studies examining the psychometric properties of the *FLACC* scale in different populations and settings. For the assessment of postoperative pain, 10 studies explored the reliability of the *FLACC* scale. Five studies revealed -near-perfect reliability in infants and children aged 0 to 10 years from PACU, intensive care unit (ICU) and inpatient unit. The remaining 5 studies showed moderate reliability for infants and children in cleft lip and palate unit, and burns/surgical/trauma unit. For the assessment of procedural pain, 4 studies tested good reliability in infants and children aged 6 months to 7 years from emergency unit, immunization center and PICU.

Due to the absence of a gold standard for pain assessment, the criterion validity of the *FLACC* scale could not be tested.[30,31] The 15 included studies all used different pain scales as reference standard to assess convergent validity, leading to the significant difference of findings. For assessment of postoperative pain, overall correlation with other pain scales were moderate to high in PACU, PICU and inpatient unit, and were high in cleft lip and palate unit, and burns/surgical/trauma unit. For procedural pain assessment, the overall correlation with other pain scales was excellent in emergency unit, immunization center and PICU.

Examining the responsiveness of the *FLACC* scales suggested that pain intensity could vary over time. Hence, it is important to perform the test-retest analysis. However, only 2 studies in China completed test-retest analysis and one reported satisfactory test-retest reliability.[14,24] Inadequate test-retest analysis would be the main reason of decreasing the overall quality of all the included studies in the present study. Hence, further studies with test-retest analysis are strongly recommended.

Some limitations in this study must be acknowledged. First and foremost, we only searched PubMed, EMBASE, OVID and CNKI, no other databases were searched. Therefore, some studies may be missed. Notably, only 4 new publications were synthesized into the present systematic review since the last systematic review was published. However, due to the searching in accordance with Cochrane handbook for performing systematic review, we still convince that the minor limitation may not significantly impair the findings. We then included only studies published in English and Chinese; therefore, some potentially eligible studies in other languages were also missed. Second, the present systematic review included different language versions of the *FLACC*, and findings related to a language version are not necessarily generalizable to the other languages version. Third, we also admitted that we performed the present systematic review only following the standard procedures the Cochrane collaboration recommended rather than referencing the methodological framework developed by the consensus-based standards selecting of health status measurement instruments (COSMIN) group. Notably, the COSMIN methodological framework has been extensively recommended as the standard for performing a systematic review of psychometric assessment. Therefore, the future systematic review should be performed based on the COSMIN framework when further studies on this topic are available. Finally, responsiveness was assessed by discriminative capability in 9 included studies rather than using standard error of measurement (SEM) and the smallest detectable change (SDC).

Conclusions

In summary, although there is no gold standard for observational pain assessment, the data to date suggest that scores on the *FLACC* capture pain behaviors similar to other observational pain measures, supporting to use the *FLACC* from the perspective of criterion validity. Therefore, the *FLACC* scale can be considered for measuring observational pain in infants and children; however, further studies are still needed to provide more robust evidence.

Data Sharing Statement

All data generated or analyzed during this study are included in this published article/as supplementary information files.

Funding

This work was supported by the scientific research project of Hunan Provincial Health Commission (20200712).

Disclosure

The authors declare that they have no competing interests for this work.

References

1. Benini F, Congedi S, Scalfaro C, Barbi E, Podestà AF. Up-date of paediatric pain management in emergency units. *Medico e Bambino*. 2019;38(1):19–27.
2. Andersen RD, Langius-Eklof A, Nakstad B, Bernklev T, Jylli L. The measurement properties of pediatric observational pain scales: a systematic review of reviews. *Int J Nurs Stud*. 2017;73:93–101. doi:10.1016/j.ijnurstu.2017.05.010
3. Alizadeh Z, Paymard A, Khalili A, Hejr H. A systematic review of pain assessment method in children. *Ann Trop Med Public Health*. 2017;10(4):847–849. doi:10.4103/ATMPH.ATMPH_155_17
4. Latina R, Di Maglie V, Sili A, et al. Valutazione del dolore nei pazienti affetti dasindrome di Down: una revisione narrativa dellaletteratura [Pain assessment in Down Syndrome patients: a narrative review of the literature]. *Prof Inferm*. 2018;71(4):209–220. doi:10.7429/pi.2018.714209. Itailan.
5. Verstraete J, Ramma L, Jelsma J. Validity and reliability testing of the Toddler and Infant (TANDI) Health Related Quality of Life instrument for very young children. *J Patient-Rep Outcomes*. 2020;4(1):94. doi:10.1186/s41687-020-00251-4
6. Tamvaki E, Efstratiou F, Nteli C, et al. Validation of the Greek Version of Comfort-B, FLACC, and BPS Scales in Critically Ill Children and Their Association with Clinical Severity. *Pain Manag Nurs*. 2020;21(5):468–475. doi:10.1016/j.pmn.2019.10.004
7. 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(3):293–297.
8. Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Paediatr Anaesth*. 2007;17(8):720–733. doi:10.1111/j.1460-9592.2007.02218.x
9. von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain*. 2007;127(1–2):140–150. doi:10.1016/j.pain.2006.08.014
10. Crellin DJ, Harrison D, Santamaria N, Babl FE. Systematic review of the Face, Legs, Activity, Cry and Consolability scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? *Pain*. 2015;156(11):2132–2151. doi:10.1097/j.pain.0000000000000305
11. Bringuier S, Picot MC, Dadure C, et al. A prospective comparison of post-surgical behavioral pain scales in preschoolers highlighting the risk of false evaluations. *Pain*. 2009;145(1–2):60–68. doi:10.1016/j.pain.2009.05.012
12. Johansson M, Kokinsky E. The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nurs Crit Care*. 2009;14(3):122–130. doi:10.1111/j.1478-5153.2009.00323.x
13. Bai J, Hsu L, Tang Y, Van Dijk M. Validation of the COMFORT Behavior Scale and the FLACC Scale for Pain Assessment in Chinese Children after Cardiac Surgery. *Pain Management Nursing*. 2012;13(1):18–26. doi:10.1016/j.pmn.2010.07.002
14. Jia YG. *The Application Study of Three Types of Pain Behavioural Observation Scales for Pain Measurement in Young Children with Burns*. Shanghai, China: Nursing department, Jinan University; 2012.
15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
16. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. *J Eval Clin Pract*. 2012;18(6):1126–1132. doi:10.1111/j.1365-2753.2011.01707.x
17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
18. Chang Q, Sha F, Chan CH, Yip PS. Validation of an abbreviated version of the Lubben Social Network Scale ("LSNS-6") and its associations with suicidality among older adults in China. *PLoS One*. 2018;13(8):e0201612. doi:10.1371/journal.pone.0201612
19. Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health*. 1997;20(3):269–274. doi:10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G
20. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge; 2013.
21. Suraseranivongse S, Santawat U, Kraiprasit K, Petcharatana S, Prakkamodom S, Muntraporn N. Cross-validation of a composite pain scale for preschool children within 24 hours of surgery. *Br J Anaesth*. 2001;87(3):400–405. doi:10.1093/bja/87.3.400
22. Manworren RC, Hynan LS. Clinical validation of FLACC: preverbal patient pain scale. *Pediatr Nurs*. 2003;29(2):140–146.
23. Willis MH, Merkel SI, Voepel-Lewis T, Malviya S. FLACC Behavioral Pain Assessment Scale: a comparison with the child's self-report. *Pediatr Nurs*. 2003;29(3):195–198.
24. Liu M, Chen LQ, Zheng JL. The appraisal of reliability and validity of the FLACC observation tool for scoring painfulness of children with cleft lip and palate. *Nurs J Chin PLA*. 2012;7A(29):20–22,26.
25. Gomez RJ, Barrowman N, Elia S, Manias E, Royle J, Harrison D. Establishing intra- and inter-rater agreement of the Face, Legs, Activity, Cry, Consolability scale for evaluating pain in toddlers during immunization. *Pain Res Management*. 2013;18(6):e124–128. doi:10.1155/2013/897104
26. Kochman A, Howell J, Sheridan M, et al. Reliability of the Faces, Legs, Activity, Cry, and Consolability Scale in Assessing Acute Pain in the Pediatric Emergency Department. *Pediatr Emerg Care*. 2017;33(1):14–17. doi:10.1097/PEC.0000000000000995

27. Crellin DJ, Harrison D, Santamaria N, Huque H, Babl FE. The Psychometric Properties of the FLACC Scale Used to Assess Procedural Pain. *J Pain*. 2018;19(8):862–872. doi:10.1016/j.jpain.2018.02.013
28. Matsuishi Y, Hoshino H, Shimojo N, et al. Verifying the validity and reliability of the Japanese version of the Face, Legs, Activity, Cry, Consolability (FLACC) Behavioral Scale. *PLoS One*. 2018;13(3):e0194094. doi:10.1371/journal.pone.0194094
29. Tsze DS, Ieni M, Flores-Sanchez PL, et al. Quantification of Pain and Distress Associated With Intranasal Midazolam Administration in Children and Evaluation of Validity of Four Observational Measures. *Pediatr Emerg Care*. 2021;37(1):e17–e20. doi:10.1097/PEC.0000000000001526
30. Karimi Z, Baraz S, Rashidi M, Jahani S, Malehi AS. Pain assessment tools in children undergoing mechanical ventilation in intensive care units. *Indian J Forensic Med Toxicol*. 2018;12(4):184–188. doi:10.5958/0973-9130.2018.00221.9
31. Eriksson M, Campbell-Yeo M. Assessment of pain in newborn infants. *Semin Fetal Neonatal Med*. 2019;24(4):101003. doi:10.1016/j.siny.2019.04.003