

## Accepted Manuscript

Title: Infant Clinical Pain Assessment: Core Behavioural Cues

Author: Miranda G. DiLorenzo, Rebecca Pillai Riddell, David B. Flora,  
Kenneth D. Craig

PII: S1526-5900(18)30140-8

DOI: <https://doi.org/10.1016/j.jpain.2018.03.016>

Reference: YJPAI 3565

To appear in: *The Journal of Pain*

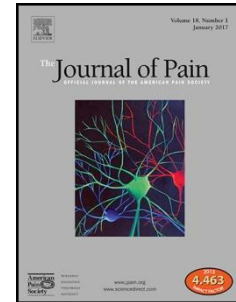
Received date: 28-7-2017

Revised date: 31-12-2017

Accepted date: 8-3-2018

Please cite this article as: Miranda G. DiLorenzo, Rebecca Pillai Riddell, David B. Flora, Kenneth D. Craig, Infant Clinical Pain Assessment: Core Behavioural Cues, *The Journal of Pain* (2018), <https://doi.org/10.1016/j.jpain.2018.03.016>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



## Infant Clinical Pain Assessment: Core Behavioural Cues

Miranda G. DiLorenzo <sup>a</sup>, Rebecca Pillai Riddell <sup>a,b,c</sup>, David B. Flora <sup>a</sup>, & Kenneth D. Craig <sup>d</sup>

<sup>a</sup> York University, Toronto, Canada

<sup>b</sup> Hospital for Sick Children, Toronto, Canada

<sup>c</sup> University of Toronto, Toronto, Canada

<sup>d</sup> University of British Columbia, Vancouver, Canada

*Corresponding author:* Dr. Rebecca Pillai Riddell. *Mailing address:* The Opportunities to Understand Childhood Hurt (OUCH) Laboratory, York University, Department of Psychology, Faculty of Health, Room 2006 Sherman Health Sciences Building, 4700 Keele Street, Toronto, ON, M3J 1P3 *Phone:* 416-736-2100 x 20177 *Fax:* 416-736-5814

*Email address:* [rpr@yorku.ca](mailto:rpr@yorku.ca)

*URL:* <http://www.yorku.ca/ouchlab/index.html>

*Disclosures:* The authors have no conflicts of interest to report. This research was funded by awards to Dr. Pillai Riddell from the Canadian Institutes of Health Research (CIHR) operating grants [524563, 524432] and New Investigator salary support, the Ontario Ministry of Research and Innovation Early Researcher Award, the Canadian Foundation for Innovation [532009], and salary support from her inaugural York Research Chair in Pain and Mental Health. Ms. DiLorenzo was awarded funds from the Social Sciences and Humanities Council of Canada, the Government of Ontario, and the Meighen Wright Foundation. Ms. DiLorenzo is also a trainee member of Pain In Child Health (PICH), a strategic research training initiative of CIHR.

**Highlights**

- NFCS and MBPS coding systems were examined using Confirmatory Factor Analysis.
- A revised 3-item NFCS maintained good psychometric properties of 7-item version.
- Redefinition of MBPS with cry as a sole indicator was suggested.
- The revised scales increase efficiency of coding based on improved psychometrics.

**Abstract**

Diverse behavioural cues have been proposed to be useful cues in infant pain assessment, but there is a paucity of evidence based on formal psychometric evaluation to establish their validity for this purpose. We aimed to examine two widely-used coding systems, the Neonatal Facial Coding System (NFCS) and the Modified Behaviour Pain Scale (MBPS), by examining their factor structures with confirmatory factor analysis using a large archival dataset. The results indicated that an item-reduced NFCS scale with three items produced a one-factor pain model that maintained the good psychometric properties of the 7-item scale. In addition, it was found that MBPS also has challenging internal consistency, with items that are both weakly correlated and highly redundant. One-item of the MBPS may be able to capture the construct of pain equally well or potentially improve its psychometric properties. Redefinition of MBPS with cry as a sole indicator was suggested. This analysis provides two new iterations of the NFCS and MBPS that improve construct validity and internal consistency. These shorter versions also improve the feasibility of both measures and increase their potential for clinical use as less time is required for their administration.

**Perspective:** This article presents new iterations of the NFCS and MBPS scales. These revised measures improve the internal consistency of the measures, feasibility of use of the tools in research settings and the efficiency of the coding process. The revised tools could also improve the feasibility of coding within clinical settings.

**Keywords:** pain assessment; behavioural pain measures; acute pain; infant

## Introduction

Pain assessment is the foundation of pain management, with self-report often considered the primary source of assessment in capable children and adults [40]. In preverbal infants, clinicians and researchers are wholly dependent on proxy judgments by benevolent adult caregivers, posing challenges with notable drawbacks [25,29]. To ‘hear’ the voice of the infant we must use behavioural measures. Raised awareness over recent decades of the prevalence and serious consequences of acute pain in infants has led to a proliferation of pain assessment tools; unfortunately, many fail to meet rigorous psychometric standards as a consequence of not using optimal strategies for evaluating validity [15,35].

The majority of infant pain assessment tools focus on infant non-vocal and cry behaviours [20,34]. This represents sensitivity to the main modalities whereby infants communicate pain to care providers. Of all pain behaviours, facial expression has emerged as the most reliable and consistent indicator of infant pain across populations and contexts due to its universality and specificity [36]. Crying and gross motor movements are also common responses to pain in infants and considered to be sensitive indicators of pain [9].

A number of coding systems are used to quantify behaviour responses to pain in infancy. Two widely-used measures are the Neonatal Facial Coding System (NFCS) [16], used to quantify infants’ facial expressions, and the Modified Behaviour Pain Scale (MBPS) [38], a measure combining facial expression, crying, and body movement in response to pain. Both scales display good psychometric properties [36]. However, a confirmation of the factor structure of MBPS has not been provided. Although a single component solution has been supported using the facial indicators of NFCS [11,21], sample sizes were limited and the factor

structure was not examined at different infant ages or at different phases of acute pain procedures. Further, principal component analysis (PCA) was used to examine the structure of NFCS rather than a procedure based on the common factor model [23]. Lastly, PCA is limited as a statistical method because it does not separate error variance from common factor variance, thereby preventing substantive interpretation of components [23,29]. Based on these limitations and using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) [24], it could be surmised that the methodological quality of the internal consistency investigations of NFCS has been limited.

Work on optimizing observational behaviour pain tools is critical given dependence on them in making important decisions in clinical pain management. The current study investigates the internal structure of NFCS and MBPS scales for acute pain in infancy. Although these scales are extensively used in research, information about their psychometric properties is incomplete. Therefore, the main goal of the study was to examine and confirm the factor structures of NFCS and MBPS by means of confirmatory factor analysis using an archival longitudinal dataset. Convergence between MBPS and NFCS scores was also examined as well as the concurrent associations between the pain scales and caregiver behaviours.

## **Method**

### ***Participants***

The proposed analysis used archival data from a large research study (the OUCH Cohort) [26] in which infant-caregiver dyads were recruited from 3 pediatric clinics in Toronto and observed throughout the infant's first year of life at their 2-, 4-, 6-, and 12-month immunization appointments and again at the child's preschool (i.e., 4-6 years of age) immunization. The data for this longitudinal cohort study was collected between October 2007 and December 2015. The

total sample size includes data for 760 infant-caregiver dyads. Infant pain data available from the 2-month ( $n=500$ ) and 12-month ( $n=548$ ) waves will be included in the analyses. Since there are important developmental changes across the first year of life that affect infant pain-related distress (e.g., the increasing variability of infant pain reactivity and regulatory responses [2]), the factor structures of the scales were examined for stability across different infant developmental stages. The two extremes of age in the first year of life, 2- and 12-months, were selected to examine the stability of the factor structure while keeping the results succinct with fewer factor analyses.

Generally, the participants were healthy, from middle-class families and culturally diverse, and had caregivers who were well-educated. Caregivers who were able to fluently speak English and had infants without suspected developmental delay and never admitted to the Neonatal Intensive Care Unit were eligible to participate in the study.

### ***Procedure***

Details of the procedures and materials used for the longitudinal OUCH cohort study have been previously described [26]. Here is a brief summary of the waves relevant to the current analysis. Ethical approval was obtained from the affiliate university and the associated tertiary-level hospital. After primary caregivers of the infants agreed to speak to a researcher about the study, research assistants explained the study and caregivers were asked to sign informed consent forms.

During the 2-month and 12- month immunization appointments, infants' facial expression, paralinguistic vocalizations, and body movements were video recorded for up to 2 minutes prior to the immunization and 5 minutes after the immunization. The OUCH cohort

study used naturalistic observation with minimal interference from the research team to ensure that spontaneous pain behavior of the infant in the immunization context was captured.

### *Measures*

#### **Parent demographic information**

Prior to each immunization appointment, caregivers completed a short demographic questionnaire inquiring about their relationship with the infant, education level, and self-reported heritage culture, as well as information related to the infant, such as infant age, gender, and medical conditions since the last time they participated in the study.

#### **Infant Behavioural Pain Measures**

*Neonatal Facial Coding System.* NFCS [16] was used to measure infants' facial responses to acute pain. Based on the use of NFCS in previous work [11,25,27], seven indicators (brow bulge, eye squeeze, nasolabial furrow, open lips, vertical stretch mouth, horizontal stretch mouth, taut tongue) were utilized to examine the structure of the scale. Each of the facial actions was coded as 0 (not present) or 1 (present) for every second within a 10-second period. As per standard protocol for the NFCS, for every 10 seconds of coding, each facial action receives a total score of 0 to 10, based on the presence or absence of the facial action for each of the 10 one-second epochs. Total scores are based on a sum of all visible facial action units over the 10-second epoch. Higher scores indicate greater pain-related distress. For the purpose of the current study, pain scores were obtained by coding the presence of facial actions during the 10-second period occurring immediately after the last needle the infant received (NFCS Immediate) and during the 10-second period 1 minute after the last needle (NFCS Post 1). Trained research assistants, blind to study hypotheses, coded the data. Twenty percent of the data was double-

coded for inter-rater reliability. Reliability was excellent for all seven facial actions, ranging from .85 to .97.

The original NFCS study [16] provided evidence of adequate psychometric properties. Strong reliability has been consistently demonstrated [36]. Adequate convergent validity has been shown by comparing the NFCS with other facial coding systems [10,21]. Evidence in support of the construct validity of NFCS has been established as the scale differentiated infants receiving or not receiving pharmacologic treatment [39]. However, although NFCS was sensitive to different levels of distress in infants over a procedure, it did not distinguish between pain-related and non-pain related distress [1]. Further, using PCA, a single pain component was reported when examining seven facial actions in preterm and newborn infants [11] and five facial actions in infants 2 to 18 months of age [21]. As previously mentioned, the factor structure was not examined across different infant ages or at different phases of the acute pain procedure. These are critical omissions given age and procedural differences have been reported [2] using the more detailed Facial Action Coding System [12].

*Modified Behaviour Pain Scale.* MBPS [37] uses three infant ordinal pain behaviour scores: facial expression (range 0-3), cry (range 0-4), and body movement (range 0-3) to represent the degree of infant pain-related distress. Higher scores reflect greater pain. As per standard protocol for the measure, behaviours observed during 2-month and 12-month immunization appointments were analyzed at two different time points: during the 15 seconds immediately following the last needle (MBPS Immediate) and one minute after the initial 15-second epoch (MBPS Post 1). Inter-rater reliability was calculated for twenty percent of the data. Intraclass correlations were high, ranging from .93 to .96.



In the immunization context, MBPS has displayed moderate to high concurrent validity with VAS ratings by pediatricians, as well as high interrater reliability [37,38]. Taddio and colleagues [37] also provided evidence of high internal consistency of MBPS using coefficient alpha during baseline ( $\alpha = .94$ ) and injection ( $\alpha = .83$ ) phases. Of note, inter-item correlations were not examined to identify possible redundancy in items that could have inflated alpha estimates. Further, the alpha coefficients were based on a small sample of 40 infants who ranged in age from 2 to 6 months. In support of the construct validity of the scale, MBPS discriminated between infants given placebo versus analgesia treatment [37]. To our knowledge, factor analyses have not been conducted on the MBPS behavioural indicators to model the assumed underlying pain construct.

### **Caregiver Behaviour Measures**

*Emotion Availability Scales-Fourth Edition (EAS)*. Caregiver sensitivity, looking at whether behaviours that effectively addressed their infant's pain were displayed, was coded at the 2- and 12- month immunizations, using the Infancy/Early Childhood Version of the EAS [3]. This rating scale requires clinical judgment of the quality of caregiver behaviours instead of a simple frequency count of behaviours. Higher scores indicate greater sensitivity to their infant's pain-related distress. The overall EAS composite score sums caregiver behaviours on four different subscales: sensitivity, structuring, non-intrusiveness, and non-hostility. Caregiver sensitivity reflects the caregiver's ability to interpret and respond to the infant's cues while displaying appropriate affect and considering the developmental capabilities of the infant (e.g. sensitively and contingently responding to infant's cues). Caregiver structuring refers to the caregiver's ability to structure the environment in a way that leads the infant in a positive direction (e.g. effectively uses toys to distract the infant from pain). Caregiver non-intrusiveness

involves the ability to be available for the child and avoid overbearing and overstimulating behaviours (e.g. intrusively kissing infant while they are highly distressed). Lastly, caregiver non-hostility refers to the caregiver's ability to abstain from antagonizing or displaying impatient behaviours (e.g. expressing annoyance or sarcasm).

The EAS total score was based on the video footage from the time the caregiver and infant entered the clinic room until they left. After reviewing the entire video, coders provided a rating on each of the EAS subscales (score ranges: 7-29). The subscale scores were then summed to form a composite score that potentially ranged from 28 to 116, with higher scores representing more optimal caregiver behaviours and interactions with child. Four coders who had received training with the scale's designer coded the footage from the immunization appointments. Interrater reliability was calculated among all permutations of coders (e.g. coder A with B, and A with C, and A with D) for twenty percent of the data. The intraclass correlations for the EAS composite score ranged from .80 to .93.

*Measure of Adult and Infant Soothing and Distress (MAISD)*. In contrast to clinical judgements about the sensitivity of parent behaviours in the EAS (the 'how'), the MAISD measured how often discrete behaviours were enacted by parents (the 'how much') [8]. The MAISD is a reliable and valid behavioral observation scale that was originally developed to evaluate the behaviors of infants, caregivers, and clinicians during pediatric medical procedures. Three of the eight MAISD caregiver behaviours (rocking, physical comfort, verbal reassurance) were included in the current analyses, as they have been shown to have the strongest relationship with infant pain-related distress [6,22]. The three caregiver behaviours were coded as present (1) or absent (0) for five-second epochs during the following three 60-second periods: (1) the one-minute period prior to the needle, (2) the one-minute period following the last needle, and (3) the

two-minute period following the last needle. Index scores, reflecting the proportion of time each behavior was present, were calculated by adding the total number of five-second epochs within each one-minute period each behavior was displayed divided by the total number of codable epochs within each minute. The index scores ranged from 0 to 1, with higher scores reflecting greater frequency of caregiver behaviours over a one minute period.

Seven coders were trained to reliability by the scale developer. Twenty percent of all data was coded for reliability. Inter-rater reliability was calculated among all permutations of coders. The intraclass correlations were strong and ranged from .75 to .95.

### *Statistical Analyses*

Because previous research has not provided clear guidance on the number of factors underlying NFCS and MBPS scales, we aimed to examine the factor structure of both scales. Analyses were planned for separate factor analyses of data from 2 and 12 months of age and from two different time points post-needle (immediately after the needle and 1-minute after the needle). Because factor analysis is a procedure by which a statistical model is developed to represent the pattern of correlations influenced by one or more constructs, we first examined the correlations among items on both NFCS and MBPS scales. Product-moment correlations between NFCS facial actions were examined. Because MBPS pain behaviour indicators are ordinal, matrices of polychoric correlations among indicators were calculated instead of product-moment correlations (see [13]). In order to continue our examination of scale structure with factor analysis, the items on each scale were expected to be moderately correlated (between .30 and .70) with each other [5,7,19], as they are presumed to be influenced by the same construct (i.e., pain).

Using confirmatory factor analysis (CFA), models were fitted to covariances among scale items using maximum likelihood estimation with robust standard errors and fit statistics, as implemented with “MLR” estimation using the *lavaan* package [32] in R [31]. Model fit was evaluated using the root mean square error of approximation (RMSEA), the comparative-fit index (CFI), and the Tucker-Lewis index (TLI). As rough guidelines, RMSEA values of 0.06 or lower indicate good fit along with CFI and TLI values of 0.95 or higher [18].

Once the final factor structures were obtained across age and post-needle time points, the internal consistency of the scale was to be explored using coefficient alpha. Further, after confirming the factor structures and determining the optimal item compositions of both scales, we planned a preliminary correlational analysis of the convergent validity with both measures. Additionally, because the association between infant pain-related distress and caregiver behaviours is one of the most reliable infant pain relationships documented in the literature [e.g.,14,22], concurrent validity between the pain measures and caregiver behaviours also was examined. Spearman correlations were used to examine the relationships between NFCS, and MBPS and caregiver behaviours because the distributions of the associations were not linear.

## **Results**

### ***An examination of the factor structure of NFCS***

#### **Relations among NFCS items**

Correlations among items on the NFCS ranged from extremely low to very high ( $r = .07$  to  $.92$ ) across age and within age across the post-immunization period (contact [ouchlab@yorku.ca](mailto:ouchlab@yorku.ca) for inter-item correlation tables). In particular, high correlations were apparent between brow bulge, eye squeeze, and nasolabial furrow ( $r = .71$  to  $.92$ ), suggesting potential redundancy among these items. This means only one rather than three items was needed to provide roughly the same amount of information about infant pain-related distress. Based on extensive experience of coding and analyzing thousands of data points using NFCS, as

well as statistics (redundancy) and pragmatics (coding ease), a consensus of the research team was sought to determine which redundant items to remove. Because nasolabial furrow automatically emerges when an infant's mouth is stretched horizontally (another item on the scale), it was not retained as part of the scale. Moreover, brow bulge was removed because it is statistically redundant and eye squeeze is more obvious and easier to code. Finally, taut tongue and open lips evidenced multiple weak associations with other NFCS facial actions ( $r = .07$  to  $.29$ ). This result suggests that taut tongue and open lips may not measure the same construct (i.e., pain) as other items on the scale. Taut tongue was removed for further analyses. Open lips was also initially removed before carrying out confirmatory factor analysis, but this resulted in just-identified models with trivially perfect fit. Thus, open lips was retained for factor analysis of NFCS to ensure the degrees of freedom of the estimated models were greater than zero.

Next, we investigated the factor structure of the four retained NFCS facial actions, including eye squeeze, vertical stretch mouth, horizontal stretch mouth, and open lips. Confirmatory factor analysis was used to test the potential unidimensional factor structure of pain underlying the four facial actions.

### **Confirmatory Factor Analyses of NFCS at 2- and 12-months**

*Model Fit.* Fit indices for a one-factor CFA model consistently indicated that the one-factor pain models had good fit to the NFCS data at all time points: a) 2-months immediately after the needle (RMSEA= 0.05, CFI= 0.99, TLI= 0.97); (b) 2-months at 1-minute post-needle (RMSEA= 0.05, CFI= 1.00, TLI= 0.99); (c) 12-months immediately after the needle (RMSEA= 0.06, CFI= 0.99, TLI= 0.97); and (d) 12-months at 1-minute post-needle (RMSEA= 0.06, CFI = 0.99, TLI= 0.97) . Based on the fit indices, the one-factor model accounts for covariation among the three NFCS facial actions across infancy and post-needle time points. Of note, models with more than one factor could not be examined because there must be a minimum of three items per factor for the model to converge to a proper, unique solution [4].

*One-factor model interpretation.* The standardized factor loadings for each one-factor solution at 2- and 12-months are presented in Table 1. The loadings for the pain factor across age and post-needle time points are consistently moderate-to-strong for three of the facial actions (eye squeeze, vertical stretch mouth and horizontal stretch mouth), ranging from .45 to .95. As expected based on the correlations among NFCS items, open lips had weak loadings across all time points (.23 - .35) compared to the other items. While initially entered into the CFA models to prevent trivial perfect fit, its removal (from further analysis) was now justified because it does not have a strong and stable relationship with the latent construct of pain.

### **Internal consistency of revised NFCS**

The internal consistency of the one-factor model with three NFCS items was determined based on coefficient alpha. The internal consistency of the 3-item NFCS was consistently good for all confirmed models: (a) 2-months immediately after the needle ( $\alpha = .75$ ); (b) 2-months at 1 minute post needle ( $\alpha = .87$ ); (c) 12-months immediately after the needle ( $\alpha = .76$ ); and (d) 12-months at 1 minute post-needle ( $\alpha = .75$ ). These coefficient alpha estimates are similar to those of the 7-item NFCS scale: (a) 2-months immediately after the needle ( $\alpha = .76$ ); (b) 2-months at 1 minute post-needle ( $\alpha = .87$ ); (c) 12-months immediately after the needle ( $\alpha = .81$ ); and (d) 12-months 1 minute post-needle ( $\alpha = .79$ ), suggesting that reliability of the scale was maintained despite item reduction.

### **Concurrent validity with caregiver behaviours in the acute pain context**

There were significant negative correlations between the 3-item NFCS total scores and EAS caregiver sensitivity across time points (see Table 2). These relationships suggest that the more sensitive the caregiver, the lower the infant pain scores, although the effects are generally small. There were also significant positive relationships between MAISD caregiver behaviours (physical comfort, rocking, and verbal reassurance) and NFCS scores, suggesting that infant distress is related to more caregiver attempts to soothe in the same phase or future phase of the

procedure, although these effects are also small. The correlations between the seven-item NFCS total score and EAS and MAISD caregiver behaviours are of similar magnitude (less than or equal to .27) as the correlations between the three-item NFCS total score and caregiver behaviours suggesting the concurrent validity of NFCS with caregiver behaviours is unchanged using a more parsimonious and construct valid structure.

### ***An examination of the factor structure of MBPS***

#### **Relationships among MBPS items**

The associations among MBPS behaviours at 2 and 12 months were variable (contact [ouchlab@yorku.ca](mailto:ouchlab@yorku.ca) for inter-item correlation tables). There were weak, non-significant inter-item correlations at 2 months between face and body movement right after the needle ( $r = .12$ ), suggesting that body movement may not be influenced by pain to the same extent as facial actions and cry. At 2 months 1-minute post-needle, all items are highly correlated (.79 - .97) and could be considered redundant. There were also very high correlations among items at 12 months (face and body movement,  $r = .90$ ; face and cry,  $r = .83$ ) at 1-minute post-needle, again suggesting item redundancy in these scale items. Removal of redundant and weakly associated items precludes examining the factor structure of MBPS because too few items would be available to fit factor models.

#### **Convergent validity of individual MBPS behaviours with NFCS (3- and 7-item version)**

The final step of the analyses examined the associations among NFCS total scores (with a total score based on 3 items) and the three individual MBPS pain behaviours to determine if certain MBPS indicators were more or less indicative of pain-related distress in infancy. All MBPS indicators were significantly related to the 3-item NFCS total scores across age and post-procedure, providing evidence of convergent validity (see Table 3). MBPS total scores were also significantly related to the 3-item NFCS total scores. The strength of the correlations between

MBPS indicators and total scores and the 7-item NFCS total scores were similar in magnitude to correlations between MBPS and the 3-item NFCS scale. This suggests that the convergent validity between the scales was maintained after the refinement of the scales. It is important to acknowledge, however, that some of the statistically significant correlations reported in Table 3 are relatively weak to moderate, particularly correlations between body movement and the 3-item NFCS total scores ( $r = .17$  to  $.39$ ).

## **Discussion**

The findings from this study provide important new psychometric information about NFCS and MBPS. This represents a fundamental advance in the literature of pediatric pain assessment, as previous studies examining the psychometric properties of NFCS and MBPS were limited [11,21]. We aimed to address these limitations by examining the factor structure of NFCS and MBPS by means of CFA.

### ***Factor structure of the (item-reduced) NFCS***

A unidimensional pain factor structure of NFCS was confirmed with three items (eye squeeze, vertical stretch mouth, horizontal stretch mouth). The internal consistency of the 3-item NFCS scale was reliably good across the four time and age combinations. The coefficient alpha estimates are similar for the 7-item and 3-item NFCS scales, suggesting that the removal of items does not reduce the internal consistency of the measure. Based on these analyses, it seems clear that eye squeeze, vertical stretch mouth, and horizontal stretch mouth carry the bulk of the information about infant acute pain and these psychometric properties are robust across the first year of life.

### ***Concurrent validity of NFCS with caregiver behaviours***

Small magnitude estimates of concurrent validity between the revised 3-item NFCS and caregiver behaviours was also evidenced in the current study. This is in line with previous research using the same dataset but the full seven-item version of the scale [26], NFCS scores showed small negative relationships with EAS caregiver sensitivity at 12 months. There was also weak concurrent validity between the 3-item NFCS pain scores and MAISD caregiver



behaviours (physical comfort, rocking and verbal reassurance). In line with previous research by Lisi et al. [22], again using the same dataset but the full 7-item version of the scale, small positive associations were found between pain scores and caregiver behaviours. The magnitude of relationships between NFCS and caregiver behaviours were similar for both the seven- and three-item versions of NFCS. This supports the use of the 3-item NFCS, as the concurrent validity of NFCS with caregiver behaviours remained the same after items were removed.

Based on our findings, we conclude that the 3-item NFCS can be used to reliably and validly measure acute pain in healthy infants over the first year of life. Given the strong psychometric properties of NFCS identified in this current study and others (summarized in [36]), facial activity is a solid choice for assessment of acute pain in infants in research settings. Moreover, a 3-item version is easier to code and may promote bedside utilization by clinicians and caregivers.

#### ***Poor Internal consistency of MBPS***

Poor internal consistency of MBPS was shown based on the correlations among MBPS items. There were very weak and excessively high inter-item correlations between items on MBPS. Weak correlations are problematic for the assumption that all items measure the same construct, that of pain. Conversely, use of highly redundant items creates an overly narrow scale of pain that does not assess the construct optimally. Weakly associated items, as well as redundant items, were considered for removal.

#### ***Convergent validity between MBPS indicators and the three-item NFCS***

Overall, the results demonstrated convergent validity between MBPS indicators and sum scores and NFCS sum scores. Generally, the strength of the associations between MBPS and NFCS were the same for both the seven- and three-item versions of NFCS and the one- and three-item versions of MBPS. Their convergences between different versions of the scale suggest that a more parsimonious version of each measure does not impact the important established relationships of these commonly used infant pain measures.

Importantly, as previously discussed, the original 7-item version of the NFCS was not reflective of a unitary construct, lending a serious threat to construct validity. Moreover, it is important to note that the majority of associations between NFCS sum scores and MBPS individual behaviours were relatively small/moderate in magnitude. Relationships were weaker particularly when pain was measured immediately after the needle and more so at 2-months of age. This suggests that NFCS and MBPS may not be measuring the same construct (i.e. pain-related distress) at this age and post-needle time point. Taking a lifespan perspective, weaker relationships between younger infant behaviours make some sense. There is a steeper developmental trajectory closer to birth, which may make it harder to demonstrate stable relationships due to constantly changing developmental capacities [33].

Furthermore, body movement on MBPS generally had the weakest association with NFCS scores. It appears that body movement is not as sensitive in measuring acute pain compared to face and cry. There was also redundancy between face and cry. Only one item of the two needs to be retained on the scale to gather the same amount of information that would be collected with both items. Extensive research experience coding thousands of infant vaccination appointments with this measure was used to determine which item to remove. Face was removed from MBPS because cry is pragmatically easier to observe and a more comprehensive indicator of pain-related distress on MBPS (i.e. it embeds in the coding of this item a baseline assessment of crying). Moreover, the MBPS cry item adjusts for pre-needle distress which increases the sensitivity of the indicator [1].

Surprisingly, the face item on MBPS had lower than expected correlations with NFCS, given they are measuring the same behavioural cue. Upon reflection, it is noted that there is a fundamental difference in how facial cues are coded on NFCS and MBPS. NFCS offers a more fine-grained analysis of different facial actions associated with pain-related distress in infants, whereas MBPS codes broader facial expressions (e.g. a score of 1 is given to infants with a “neutral expression”). Considering the relationships among MBPS items, the associations with NFCS total scores, and using our extensive experience coding MBPS within the immunization

context, we propose removing body movement and face from MBPS to improve the reliability and validity of the scale.

### ***Implications for Scale Use and Future Directions***

The current findings provide important implications for new versions of the MBPS and NFCS to assess acute pain in infants. First, with improved construct validity and internal consistency, both scales are more accurate in capturing a unidimensional construct of pain (which is how it has been traditionally used without a confirmatory factor analysis), and consequently, will have better diagnostic value than the longer versions of the scales. Second, the results of this study support the use of the 3-item NFCS as an instrument to assess acute pain in preverbal infants, particularly for research purposes, but item reduction improves its potential for use in clinical settings. With ten or seven facial actions to code, the training and reliability coding process can be challenging and lengthy. The revised NFCS with three items improves the feasibility of use of the tool in research settings and the efficiency of the coding process. The revised NFCS could also improve the feasibility of use in clinical settings because it seems more feasible to quickly discern eyes and mouth in vivo. Future research should be conducted in clinical practice using the three-item NFCS to assess the feasibility and clinical utility of the tool with health professionals and parents using a similar method of application as Grunau and colleagues [17] who found the 10-item NFCS feasible to code in real-time at bedside.

Finally, MBPS may not produce the most valid assessment of infant acute pain in its current state. A revision of MBPS, with only cry as an indicator, is suggested for use to assess infant pain-related distress in acute pain contexts over the first year of life. Similar to NFCS, using cry as a sole indicator on MBPS should be examined in clinical settings to determine its feasibility and clinical utility. In addition, more research is needed to examine the sensitivity of both revised measures and whether these tools can measure pain in other clinical populations (e.g., premature infants, critically ill neonates).

Considering all the results of the current study together, new recommendations can be made regarding behavioural assessment of pain related distress in infancy. If time permits the use

of NFCS with three facial actions should be used because these three items can provide greater sensitivity to the variability of infant reactions without the challenges of redundancy. However, if facial activity cannot be coded, assessing cry, using the MBPS operationalization, can also provide a more efficient way to accurately assess infant pain-related distress. For the broader literature, this article sets a precedent for future research by highlighting the importance of using psychometric rigor to optimize the scales used to assess infant pain that directly inform pain management.

### **Acknowledgments**

The authors would like to gratefully acknowledge their funding sources. This research was funded by awards to Dr. Pillai Riddell from the Canadian Institutes of Health Research (CIHR) operating grants and New Investigator salary support, the Ontario Ministry of Research and Innovation Early Researcher Award, the Canadian Foundation for Innovation, and salary support from her inaugural York Research Chair in Pain and Mental Health. Ms. DiLorenzo was awarded funds from the Social Sciences and Humanities Council of Canada, the Government of Ontario, and the Meighen Wright Foundation. Ms. DiLorenzo is also a trainee member of Pain In Child Health (PICH), a strategic research training initiative of CIHR.

### References

- [1] Ahola Kohut S, Pillai Riddell R. Does the neonatal facial coding system differentiate between infants experiencing pain-related and non-pain-related distress?. *J Pain* 10:214-220, 2009
- [2] Ahola Kohut S, Pillai Riddell R, Flora DB, Oster H. A longitudinal analysis of the development of infant facial expressions in response to acute pain: Immediate and regulatory expressions. *Pain* 153:2458-2465, 2012
- [3] Biringen Z. *The Emotional Availability (EA) Scales*, 4<sup>th</sup> ed. 2008
- [4] Bollen KA. *Structural equations with latent variables*. New York, Wiley, 1989
- [5] Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?. *Personal Individ Differ* 12:291-294, 1991
- [6] Campbell L, Pillai Riddell R, Garfield H, Greenberg S. A cross-sectional examination of the relationships between caregiver proximal soothing and infant pain over the first year of life. *Pain* 154:813-823, 2013
- [7] Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychol Assessment* 7:309, 1995
- [8] Cohen LL, Bernard RS, McClelland CB, MacLaren JE. Assessing medical room behavior during infants' painful procedures: The measure of adult and infant soothing and distress (MAISD). *Children's Health Care* 34:81-94, 2005
- [9] Cong X, McGrath JM, Cusson RM, Zhang D. Pain assessment and measurement in neonates: an updated review. *Adv Neonatal Care* 13:379-395, 2013
- [10] Craig KD, Hadjistavropoulos HD, Grunau RV, Whitfield MF. A comparison of two measures of facial activity during pain in the newborn child. *J Pediatr Psychol* 19:305-318, 1994

- [11] Craig KD, Whitfield MF, Grunau RV, Linton J, Hadjistavropoulos HD. Pain in the preterm neonate: behavioural and physiological indices. *Pain* 52:287-299, 1993
- [12] Ekman P, Friesen W. Investigator's guide to the Facial Action Coding System. Palo Alto, Consulting Psychologists Press, 1978
- [13] Flora DB, LaBrish C, Chalmers RP. Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Front Psychol* 3, 2012
- [14] Frank NC, Blount RL, Smith AJ, Manimala MR, Martin, JK. (1995). Parent and staff behavior, previous child medical experience, and maternal anxiety as they relate to child procedural distress and coping. *J Pediatr Psychol* 20:277-289, 1995
- [15] Gélinas C, Loisel CG, LeMay S, Ranger M, Bouchard E, McCormack D. Theoretical, psychometric, and pragmatic issues in pain measurement. *Pain Manage Nurs* 9:120-130, 2008
- [16] Grunau RV, Craig KD. (1987). Pain expression in neonates: facial action and cry. *Pain* 28:395-410, 1987
- [17] Grunau RE, Oberlander T, Holsti L, Whitfield MF. Bedside application of the neonatal facial coding system in pain assessment of premature infants. *Pain*, 76:277-286, 1998
- [18] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 6:1-55, 1999
- [19] Johnston CC: Psychometric issues in the measurement of pain. In: Finley GA, McGrath PJ. (eds): *Measurement of pain in infants and children*. Seattle, IASP Press, 1998, pp 5-20
- [20] Johnston C, Campbell-Yeo M, Fernandes A, Ranger M: Neonatal Pain. In: *Comprehensive Treatment of Chronic Pain by Medical, Interventional, and Integrative Approaches*. New York, Springer, 2013, pp 1015-1029

- [21] Lilley CM, Craig KD, Grunau RE. The expression of pain in infants and toddlers: developmental changes in facial action. *Pain* 72:161-170, 1997
- [22] Lisi D, Campbell L, Pillai Riddell RP, Garfield H, Greenberg S. Naturalistic parental pain management during vaccinations during the first year of life: Observational norms from the OUCH cohort. *Pain* 154:1245-1253, 2013
- [23] MacCallum RC: Factor analysis. In: Millsap RE, Maydeu-Olivares A (eds): *The SAGE Handbook of Quantitative Methods in Psychology*. Thousand Oaks, SAGE, 2009, pp 123-147
- [24] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, de Vet HC. COSMIN checklist manual. Amsterdam, University Medical Center, 2012
- [25] Oberlander TF, Grunau RE, Whitfield MF, Fitzgerald C, Pitfield S, Saul JP. Biobehavioral pain responses in former extremely low birth weight infants at four months' corrected age. *Pediatrics* 105:e6-e6, 2000
- [26] Pillai Riddell R, Campbell L, Flora DB, Racine N, Osmun LD, Garfield H, Greenberg S. The relationship between caregiver sensitivity and infant pain behaviors across the first year of life. *Pain* 152:2819-2826, 2011
- [27] Pillai Riddell RR, Craig KD. Judgments of infant pain: The impact of caregiver identity and infant age. *J Pediatr Psychol* 32:501-511, 2007
- [28] Pillai Riddell R, Fitzgerald M, Slater R, Stevens B, Johnston C, Campbell-Yeo M. Using only behaviours to assess infant pain: a painful compromise?. *Pain* 157:1579-1580, 2016
- [29] Pillai Riddell R, Flora DB, Stevens S, Greenberg S, Garfield H. The role of infant pain behaviour in predicting parent pain ratings. *Pain Res Manage* 19:e124-e132, 2014
- [30] Preacher KJ, MacCallum RC. Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics* 2:3-43, 2003

- [31] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013
- [32] Rosseel Y. lavaan: An R package for structural equation modeling. *J Stat Softw* 48:1–36, 2012
- [33] Sameroff, A. A unified theory of development: A dialectic integration of nature and nurture. *Child Dev*, 81:6-22, 2010
- [34] Sekhon KK, Fashler SR, Versloot J, Lee S, Craig KD. Children’s Behavioral Pain Cues: Implicit Automaticity and Control Dimensions in Observational Measures. *Pain Res Manage*, 2017
- [35] Stevens B, Franck L, Gibbins S, McGrath PJ, Dupuis A, Yamada J. Determining the structure of acute pain responses in vulnerable neonates. *Can J Nurs Res* 39:32-47, 2007
- [36] Stevens BJ, Pillai Riddell RR, Oberlander TE, Gibbins S. Assessment of pain in neonates and infants. In: *Pain in neonates and infants*. Edinburgh: Elsevier Unlimited, 2007
- [37] Taddio A, Hogan ME, Moyer P, Girgis A, Gerges S, Wang L, Ipp M. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing vaccination injections. *Vaccine* 29:1390-1394, 2011
- [38] Taddio A, Nulman I, Koren BS, Stevens B, Koren G. A revised measure of acute pain in infants. *J Pain Symptom Manage* 10:456-463, 1995
- [39] Taddio A, Stevens B, Craig K, Rastogi P, Ben-David S, Shennan A, Mulligan P, Koren G. Efficacy and safety of lidocaine–prilocaine cream for pain during circumcision. *New Engl J Med* 336:1197-1201, 1997
- [40] Tsze DS, von Baeyer CL, Bulloch B, Dayan PS. Validation of self-report pain scales in children. *Pediatrics* 132:e971-e979, 2013



## Table Legend

**Table 1.** Standardized factor loadings of the one-factor solution at 2 and 12 months.

NFCS Items	2 months		12 months	
	Pain Factor at NFCS Immediate	Pain Factor at NFCS Post 1	Pain Factor at NFCS Immediate	Pain Factor at NFCS Post 1
Eye Squeeze	0.45	0.74	0.73	0.71
Vertical Stretch Mouth	0.76	0.83	0.57	0.48
Horizontal Stretch Mouth	0.89	0.94	0.82	0.95
Open lips	0.23	0.23	0.35	0.22

Accepted Manuscript

**Table 2.** Spearman correlations between the 3-item and 7-item NFCS total scores and caregiver behaviours at 2 and 12 months.

	2 months				12 months			
	3-item NFCS Immediate Total Score	3-item NFCS Post 1 Total Score	7-item NFCS Immediate Total Score	7-item NFCS Post 1 Total Score	3-item NFCS Immediate Total Score	3-item NFCS Post 1 Total Score	7-item NFCS Immediate Total Score	7-item NFCS Post 1 Total Score
<b>EAS:</b>								
Caregiver Sensitivity Total	-.04	-.16*	-.03	-.13*	-.14*	-.17*	-.06	-.15*
<b>MAISD One-Minute Pre-Needle:</b>								
Caregiver Physical Comfort	.01	.03	.02	.02	-.06	-.05	.08	.05
Caregiver Rocking	.02	.05	.00	.03	-.01	.10	.03	.17*
Caregiver Verbal Reassurance	.01	-.03	.03	-.05	-.00	-.01	-.02	.05
<b>MAISD One-Minute Post-Needle:</b>								
Caregiver Physical Comfort	-.01	.05	.02	.11*	.21*	.01	.22*	.09
Caregiver Rocking	.07	.05	.08	.14*	.17*	.13*	.18*	.18*
Caregiver Verbal Reassurance	-.02	.06	.01	.15*	.09	.10	.09	.14*
<b>MAISD Two-Minutes Post-Needle:</b>								
Caregiver Physical Comfort	.07	.11*	.10*	.16*	.17*	.14*	.17*	.18*
Caregiver Rocking	.12*	.14*	.09	.25*	.27*	.19*	.27*	.19*
Caregiver Verbal Reassurance	.02	.21*	.01	.27*	.25*	.18*	.27*	.26*

Note: \* $p < .0$

**Table 3.** Spearman correlations between NFCS total scores and MBPS behaviours and total scores at 2 and 12 months.

MBPS Immediate Behaviours	2-months		12-months	
	3-item NFCS Immediate Total Score	7-item NFCS Immediate Total Score	3-item NFCS Immediate Total Score	7-item NFCS Immediate Total Score
Face	.18*	.18*	.44*	.46*
Cry	.24*	.27*	.33*	.33*
Body Movement	.17*	.15*	.28*	.26*
MBPS Total Score	.12*	.18*	.32*	.30*
MBPS Post 1 Behaviours	3-item NFCS Post 1 Total Score	7-item NFCS Post 1 Total Score	3-item NFCS Post 1 Total Score	7-item NFCS Post 1 Total Score
Face	.59*	.62*	.49*	.58*
Cry	.60*	.61*	.47*	.56*
Body Movement	.43*	.43*	.39*	.50*
MBPS Total Score	.59*	.61*	.48*	.57*

Note: \* $p < .05$