AN EXAMINATION OF THE FACTOR STRUCTURE OF THE NEONATAL FACIAL CODING SYSTEM AND THE MODIFIED BEHAVIOUR PAIN SCALE

MIRANDA G. DILORENZO

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN CLINICAL DEVELOPMENTAL PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

August 2017

Abstract

Diverse behavioural cues have been proposed to be useful indicators of infant pain, 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 pain scales, the Neonatal Facial Coding System (NFCS) and the Modified Behaviour Pain Scale (MBPS), by examining their factor structures with factor analysis. 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 weakly correlated and highly redundant. Redefinition of MBPS with a single indicator was suggested. This analysis provides new iterations of NFCS and MBPS that improve construct validity and internal consistency. These versions also increase the feasibility of both measures and improve their potential for clinical use.

Acknowledgements

This thesis is dedicated to all those who have offered their time and support to this current project over the past two years. First, I would like to thank my supervisor, Dr. Rebecca Pillai Riddell, whose mentorship, guidance and support have been imperative to my development as a graduate student.

Many thanks are due to Dr. Kenneth Craig for his brilliant insights, support and taking the time to review my thesis every step of the way. To the other members of my committee (Dr. Janessa Drake and Dr. Melody Wiseheart), thank you for agreeing to be a part of my committee and offering many thoughtful comments. I would also like to extend a sincere thank you to Dr. David Flora for generously sharing his time and statistical knowledge with me. Thank you for being patient and always offering words of encouragement through all the ups and downs.

To my brilliant lab mates (Dr. Jodi Martin, Lauren Campbell, Monica O'Neill, Jordana Waxman, Hannah Gennis and Shaylea Badovinac), thank you for all your guidance, support and encouragement, and most importantly, for the smiles and laughs over the past 2 years. I feel very fortunate to have the opportunity to work with all of you.

Last, but definitely not least, I want to thank my family and friends. To my parents, Marco and Monika, and my brother, Matthew, thank you for your never-ending love and support and encouraging me to always follow my dreams. Thank you to my grandmother, Mary, for being a constant source of inspiration. To my partner, Paul, thank you for always believing in me and encouraging me throughout this entire journey. I am eternally grateful for all your love and support.

TABLE OF CONTENTS

LIST OF TABLES

1. Introduction

Pain assessment is the foundation of pain management, with self-report often considered the primary source of assessment in capable children and adults (Tsze, von Baeyer, Bulloch, & Dayan, 2009). In preverbal infants, clinicians and researchers are wholly dependent on proxy judgments by benevolent adult caregivers, posing challenges with notable drawbacks (Mamedova, Pillai Riddell, & Flora, under review; Pillai Riddell & Craig, 2007; Pillai Riddell, Flora, Stevens, Greenberg, & Garfield, 2014). 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 (Gélinas et al., 2008; Stevens, Franck, et al., 2007).

The majority of infant pain assessment tools focus on infant non-vocal and cry behaviours (Johnston, Campbell-Yeo, Fernandes, & Ranger, 2013; Sekhon, Fashler, Versloot, Lee, & Craig, 2017). 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 (Stevens, Pillai Riddell, Oberlander, & Gibbins, 2007). Crying and gross motor movements are also common responses to pain in infants and considered to be sensitive indicators of pain (Cong, McGrath, Cusson, & Zhang, 2013).

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; Grunau & Craig, 1987), used to quantify infants' facial expressions, and the Modified Behaviour Pain Scale (MBPS; Taddio et al., 1995), a measure combining facial expression, crying, and body movement in response to pain. Both scales display good psychometric properties (Stevens, Pillai Riddell, et al., 2007), but they seem to include items that may be redundant and of questionable item validity. A confirmation of the factor structure of MBPS, to establish that the covariation among observed behaviours can be explained by one or more meaningful constructs (i.e., *pain*), has not been provided. Although a single component solution has been supported using the facial indicators of NFCS (Craig, Whitfield, Grunau, Linton, & Hadjistavropoulos, 1993; Lilley, Craig, & Grunau, 1997), 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 (see MacCallum, 2009). 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 (Preacher & MacCallum, 2003; MacCallum, 2009). Based on these limitations and using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN; Mokkink et al., 2012), it could be surmised that the methodological quality of the internal consistency investigations of the of NFCS has been limited.

Although there has been criticism regarding reliance on pain behaviours to measure pain in preverbal infants (Pillai Riddell et al., 2016), work on optimizing observational behaviour tools is critical given dependence on them in making important decisions in clinical pain management. The current study investigates the internal structure of two well-used behavioural scales for acute pain in infancy, NFCS and MBPS. 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. Internal consistency was also examined to summarize the overall strength of the associations between the items and pain construct. Finally, convergence between MBPS and NFCS scores was also examined as well as the associations between the pain scales and concurrently measured caregiver behaviours.

2. Method

2.1 Participants

The proposed analysis used archival data from a large research study (the OUCH Cohort; Pillai Riddell et al., 2016) 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. 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.

2.2 Procedure

Details of the procedures and materials used for the longitudinal OUCH cohort study are provided in Appendix A. 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. Before each immunization appointment, caregivers filled out a short demographic questionnaire. 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.

2.3 Measures

2.3.1 Infant Behavioural Pain Measures

Neonatal Facial Coding System. NFCS (Grunau & Craig, 1987) was used to measure infants' facial responses to acute pain. Based on the use of NFCS in previous work (Craig et al., 1993; Pillai Riddell & Craig, 2007; Oberlander et al., 2000), 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 (see Appendix B). 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 of absence of the facial action for each of the 10 one-second epochs. 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. Percentage agreement was high for all seven facial actions, ranging from .85 to .97.

The original NFCS study (Grunau & Craig, 1987) provided evidence of adequate psychometric properties. Strong reliability has been consistently demonstrated (Stevens, Pillai Riddell, et al., 2007). Adequate convergent validity has been shown by comparing the NFCS with other facial coding systems (Craig, Hadjistavropoulos, Grunau, & Whitfield, 1994; Lilley, Craig, & Grunau, 1997). Evidence in support of the construct validity of NFCS has been established as the scale differentiated infants receiving or not receiving pharmacologic treatment (Taddio et al., 1997). 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 (Ahola Kohut, & Pillai Riddell, 2009). Further, using PCA, a single pain component was reported when examining seven facial actions in preterm and newborn infants (Craig et al., 1993) and five facial actions in infants 2 to 18 months of age (Lilley et al., 1997). As previously mentioned, the factor structure was not examined across different infant ages or at different phases of the acute pain procedure, critical omissions given age and procedural differences have been reported (Ahola Kohut, Pillai Riddell, Flora, & Oster, 2012) using the more detailed Facial Action Coding System (Ekman & Friesen, 1978).

Modified Behaviour Pain Scale. MBPS (Taddio et al., 1995) 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 (see Appendix C). 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 between coders was high (intraclass correlations ranged 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 (Taddio et al., 2011; Taddio et al., 1995). Taddio and colleagues (2011) 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 (Taddio et al., 2011). To our knowledge, factor analyses have not been conducted on the MBPS behavioural indicators to model the assumed underlying pain construct.

2.3.2 *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 (Biringen, 2008). 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 (see Appendix D). 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). The intraclass correlations for the EAS composite score ranged from .80 to .93.

Measure of Adult and Infant Soothing and Distress (MAISD; Cohen, Bernard, McClelland, & MacLaren, 2005). 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'). 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; see Appendix E) were included in the current analyses, as they have been shown to have the strongest relationship with infant pain-related distress (Campbell, Pillai Riddell, Garfield, & Greenberg, 2013; Lisi, Campbell, Pillai Riddell, Garfield, & Greenberg, 2013). 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 behaviour 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 ranged from .75 to .95.

2.4 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 Flora, LaBrish, & Chalmers, 2012). 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 (Clark & Watson, 1995; Johnston, 1998), as they are presumed to be influenced by the same construct (i.e., pain).

The 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 (Rosseel, 2016) in R (R Core Team, 2013). 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 (Hu & Bentler, 1999).

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. Lisi et al., 2013;), concurrent validity between the pain measures and caregiver behaviours also was examined.

3. Results

3.1 An examination of the factor structure of NFCS

3.1.1 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 (see Tables 1 to 4). 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.

Table 1. Correlations among NFCS facial actions at 2 months immediately after the needle.

NFCS Facial Actions	1	2	3	4	5	6	7
1. Brow Bulge	1						
2. Eye Squeeze	.81	1					
3. Nasolabial Furrow	.74	.63	1				
4. Open lips	.24	.23	.25	1			
5. Vertical Stretch Mouth	.36	.33	.39	.17	1		
6. Horizontal Stretch Mouth	.42	.40	.53	.18	.69	1	
7. Taut Tongue	.23	.17	.25	.13	.39	.36	1

Table 2. Correlations among NFCS facial actions at 2 months, 1 minute post-needle.

NFCS Facial Actions	1	2	3	4	5	6	7
1. Brow Bulge	1						
2. Eye Squeeze	.74	1					
3. Nasolabial Furrow	.84	.71	1				
4. Open lips	.24	.22	.24	1			
5. Vertical Stretch Mouth	.53	.62	.54	.18	1		
6. Horizontal Stretch Mouth	.61	.70	.62	.21	.78	1	
7. Taut Tongue	.33	.21	.36	.14	.43	.47	1

Table 3. Correlations among NFCS facial actions at 12 months immediately after the needle.

NFCS Facial Actions	1	2	3	4	5	6	7
1. Brow Bulge	1						
2. Eye Squeeze	.63	1					
3. Nasolabial Furrow	.92	.65	1				
4. Open lips	.44	.32	.44	1			
5. Vertical Stretch Mouth	.38	.40	.40	.21	1		
6. Horizontal Stretch Mouth	.56	.61	.56	.28	.49	1	
7. Taut Tongue	.24	.20	.24	.14	.40	.29	1

Table 4. Correlations among NFCS facial actions at 12 months, 1 minute post-needle.

NFCS Facial Actions	1	2	3	4	5	6	7
1. Brow Bulge	1						
2. Eye Squeeze	.61	1					
3. Nasolabial Furrow	.86	.59	1				
4 .Open lips	.33	.21	.31	1			
5 .Vertical Stretch Mouth	.34	.29	.36	.11	1		
6 .Horizontal Stretch Mouth	.53	.66	.56	.18	.45	1	
7. Taut Tongue	.20	.14	.22	.07	.50	.31	1

3.1.2 Confirmatory Factor Analyses of NFCS at 2- and 12-months

Model Fit. Fit indices for a one-factor CFA model examined at 2- and 12-months and across post-needle time points appear in Table 5. 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 (Bollen, 1989). The fit statistics consistently indicated that the one-factor pain models had good fit to the NFCS data at all time points. 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.

One-factor model interpretation. The standardized factor loadings for each one-factor solution at 2- and 12-months are presented in Tables 6 and 7, respectively. 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.

Table 5. Goodness-of-fit summary for the one-factor pain model tested on NFCS items at 2- and 12-months and across post-needle time points.

One-Factor Pain Model	RMSEA	CFI	TLI
2-months NFCS Immediate	0.05	0.99	0.97
2-months NFCS Post 1	0.05	1.00	0.99
12-months NFCS Immediate	0.06	0.99	0.97
12-months NFCS Post 1	0.06	0.99	0.97

RMSEA Root Mean Square Error of Approximation; *CFI* Comparative Fit Index; *TLI* Tucker Lewis Index

Table 6. Standardized factor loadings of the one-factor solution at 2-months.

NFCS Items	Pain Factor at NFCS Immediate	Pain Factor at NFCS Post 1
Eye Squeeze	0.45	0.74
Vertical Stretch Mouth	0.76	0.83
Horizontal Stretch Mouth	0.89	0.94
Open lips	0.23	0.23

Table 7. Standardized factor loadings of the one-factor solution at 12-months.

NFCS Items	Pain Factor at NFCS Immediate	Pain Factor at NFCS Post 1
Eye Squeeze	0.73	0.71
Vertical Stretch Mouth	0.57	0.48
Horizontal Stretch Mouth	0.82	0.95
Open lips	0.35	0.22

3.1.3 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 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.

3.1.4 Concurrent validity with caregiver behaviours in the acute pain context

There were significant negative correlations between the 3-item NFCS total scores and caregiver sensitivity across time points (see Tables 8 and 9). 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 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 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.

Table 8. Correlations between the three-item and seven-item NFCS total scores and caregiver behaviours at 2-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
Caregiver Sensitivity Total Score	-.04	-.16*	-.03	-.13*
Caregiver Physical Comfort (One-Minute Pre-Needle)	.01	.03	.02	.02
Caregiver Rocking (One-Minute Pre-Needle)	.02	.05	.00	.03
Caregiver Verbal Reassurance (One-Minute Pre-Needle)	.01	-.03	.03	-.05
Caregiver Physical Comfort (One-Minute Post-Needle)	-.01	.05	.02	.11*
Caregiver Rocking (One-Minute Post-Needle)	.07	.05	.08	.14*
Caregiver Verbal Reassurance (One-Minute Post-Needle)	-.02	.06	.01	.15*
Caregiver Physical Comfort (Two-Minutes Post-Needle)	.07	.11*	.10*	.16*
Caregiver Rocking (Two-Minutes Post-Needle)	.12*	.14*	.09	.25*
Caregiver Verbal Reassurance (Two-Minutes Post-Needle)	.02	.21*	.01	.27*

**p < .05*

Table 9. Correlations between the three-item and seven-item NFCS total scores and caregiver behaviours at 12-months.

	3-item NFCS Immediate Total Score	3-item NFCS Post 1Total Score	7-item NFCS Immediate Total Score	7-item NFCS Post 1 Total Score
Caregiver Sensitivity Total Score	-.14*	-.17*	-.06	-.15*
Caregiver Physical Comfort (One-Minute Pre-Needle)	-.06	-.05	.08	.05
Caregiver Rocking (One-Minute Pre-Needle)	-.01	.10	.03	.17*
Caregiver Verbal Reassurance (One-Minute Pre-Needle)	-.00	-.01	-.02	.05
Caregiver Physical Comfort (One-Minute Post-Needle)	.21*	.01	.22*	.09
Caregiver Rocking (One-Minute Post-Needle)	.17*	.13*	.18*	.18*
Caregiver Verbal Reassurance (One-Minute Post-Needle)	.09	.10	.09	.14*
Caregiver Physical Comfort (Two-Minutes Post-Needle)	.17*	.14*	.17*	.18*
Caregiver Rocking (Two-Minutes Post-Needle)	.27*	.19*	.27*	.19*
Caregiver Verbal Reassurance (Two-Minutes Post-Needle)	.25*	.18*	.27*	.26*

*$p < .05$

3.2. An examination of the factor structure of MBPS

3.2.1 Relationships among MBPS items

The associations among MBPS behaviours at 2 and 12 months were variable (see Tables 10 and 11). 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.

Table 10. Interrelationships among MBPS behaviours immediately after the needle and 1-minute post-needle at 2-months.

	MBPS Immediate			MBPS Post 1		
	1	2	3	1	2	3
1. Face	1			1		
2. Cry	.38	1		.97	1	
3. Body Movement	.12	.45	1	.79	.82	1

Table 11. Interrelationships among MBPS behaviours immediately after the needle and 1-minute post-needle at 12-months.

	MBPS Immediate			MBPS Post 1		
	1	2	3	1	2	3
1. Face	1			1		
2 .Cry	.54	1		.83	1	
3. Body Movement	.73	.47	1	.90	.71	1

3.2.2 Convergent validity of MBPS with NFCS (3- and 7-item version)

The final step of the analyses examined the associations among NFCS total scores (based on the analyses in 3.1, 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 Tables 12 and 13). 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 of NFCS was maintained after the refinement of the scale. It is important to acknowledge, however, that some of the statistically significant correlations reported in Tables 12 and 13 are relatively moderate, particularly correlations between body movement and the 3-item NFCS total scores (r = .17 to .39).

Table 12. Correlations between NFCS total scores and MBPS behaviours and total scores at 2-months.

MBPS Immediate Behaviours	3-item NFCS Immediate Total Score	7-item NFCS Immediate Total Score
Face	.18*	18*
Cry	.24*	.27*
Body Movement	.17*	.15*
MBPS Total Score	.12*	.18*
MBPS Post 1 Behaviours	**3-item NFCS Post 1 Total Score**	**7-item NFCS Post 1 Total Score**
Face	.59*	.62*
Cry	.60*	.61*
Body Movement	.43*	.43*
MBPS Total Score	.59*	.61*

Note: *$p < .05$

Table 13. Correlations between NFCS total scores and MBPS behaviours and total scores at 12-months.

MBPS Immediate Behaviours	3-item NFCS Immediate Total Score	7-item NFCS Immediate Total Score
Face	.44*	.46*
Cry	.33*	.33*
Body Movement	.28*	.26*
MBPS Total Score	.32*	.30*
MBPS Post 1 Behaviours	**3-item NFCS Post 1 Total Score**	**7-item NFCS Post 1 Total Score**
Face	.49*	.58*
Cry	.47*	.56*
Body Movement	.39*	.50*
MBPS Total Score	.48*	.57*

Note: * $p < .05$

4. Discussion

The findings from this study provide important new psychometric information about NFCS and MBPS. There was warranted removal of weakly associated and redundant items on both scales. The underlying unidimensional factor structure of the revised NFCS had three facial indicators. There was acceptable internal consistency of the revised three-item NFCS and concurrent validity with caregiver behaviours. Convergent validity was also demonstrated between the revised three-item NFCS and MBPS (single items and total). Finally, there was a need for a potential reduction of MBPS items. All of these insights are specific to use of NFCS and MBPS in the acute pain context and over the first year of life.

Factor structure of the (item-reduced) NFCS

Upon examination of NFCS inter-item relationships, the magnitude of associations ranged from low to high. There is redundancy between the upper facial actions (brow bulge, eye squeeze, and nasolabial furrow). Only one of the items is needed to obtain the same amount of information that would be gathered with all three items. Using clinically informed judgment, eye squeeze was retained because it is easier for raters to observe compared to brow bulge, and nasolabial furrow appears anatomically redundant with another item on the scale (i.e., horizontal stretch mouth). Additionally, open lips and taut tongue had numerous weak associations with other scale items. These findings suggest that both open lips and taut tongue are not influenced by the same underlying psychophysiological process among infants in the first year of life as other items on the scale. But removing both items from the scale before confirmatory factor analysis would result in saturated models that yield perfect fit indices that do not test the veracity of the NFCS unidimensional factor structure. Thus, open lips was retained for confirmatory factor analysis to ensure the degrees of freedom of the estimated models were greater than zero.

Next, the factor structure of the four items (eye squeeze, vertical stretch mouth, horizontal stretch mouth, and open lips) retained on NFCS across age and over the post-needle period was examined. The unidimensional pain factor structure of NFCS was confirmed. The loadings for eye squeeze, horizontal stretch mouth and vertical stretch mouth on the pain factor were consistently strong. In line with the correlational analyses among NFCS items, open lips had weak factor loadings across post-needle time points and across age compared to the other facial actions. Consequently, the open lips item was dropped. Without open lips, the internal consistency of the scale was reliably good across the four time and age combinations. 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.

The unidimensionality of the three-item NFCS is similar to that identified in previous research with preterm and newborn infants, and infants across the first year of life. However, different items were specified to be a part of the structure. Craig and colleagues (1993) conducted principal component analyses on the seven NFCS items included in our preliminary analyses. The analysis yielded a single component solution for NFCS variables used to measure acute pain in preterm and newborn infants. Lilley and colleagues (1997) found a similar structure with five NFCS items (brow bulge, eye squeeze, nasolabial furrow, open lips and taut tongue) in infants from 2- to 18-months of age. Vertical stretch mouth and horizontal stretch mouth were not included, as they were infrequent in their study. We likely identified a slightly different factor structure with different items due to several reasons. First, as previously mentioned, both studies used PCA as the statistical method to examine the dimensionality of the data. Since PCA does not distinguish common and unique sources of variance (including random error), the components identified by PCA are conceptually and statistically different than factors identified via factor analysis (Preacher & MacCallum, 2003). Factor analysis is a better method to identify constructs that explain correlations among items as accurately as possible by explicitly distinguishing common and error variance. Second, the analyses of the Craig et al. (1993) and Lilley et al. (1997) were limited by the small sample sizes (N = 56 and N = 75, respectively). Third, Lilley et al. (1997) collapsed the data together for all ages (2, 4, 6, 12 and 18 months) due to limited sample sizes and were unable to examine the structure of the data within each group. Additionally, Peters and colleagues (2003) identified a unidimensional structure of NFCS with five items (brown bulge, eye squeeze, nasolabial furrow, horizontal stretch mouth, and taut tongue). However, this study examined NFCS in the context of assessing postoperative pain. Vertical stretch mouth did not cluster together with the five facial items in this study, suggesting that it may be a more sensitive indicator in an acute pain immunization context.

Small magnitude estimates of concurrent validity between the revised three-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 (Pillai Riddell et al., 2011), NFCS scores showed small negative relationships with caregiver sensitivity at 12 months. There was also weak concurrent validity between the three-item NFCS pain scores and caregiver behaviours (physical comfort, rocking and verbal reassurance). In line with previous research by Lisi and colleagues (2013), again using the same dataset but the full seven-item version of the scale, small positive associations were found between pain scores and caregiver behaviours, suggesting that infant distress elicits caregiving soothing behaviours during the same phase or future phases of the procedure. The correlations between the three-item NFCS and caregiver behaviours are of similar magnitude as the correlations between the seven-item NFCS and caregiver behaviours. This supports the use of the three-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 a short version of the NFCS with three items—including eye squeeze, horizontal stretch mouth, and vertical stretch mouth—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 Stevens, Pillai Riddell, et al., 2007), facial activity is a solid choice for assessment of acute pain related distress in infants in research settings. Moreover, a 3-item version is easier to code and may promote bedside utilization by clinicians and other caregivers.

Poor Internal consistency of MBPS

Poor internal consistency of MBPS was shown based on the correlations among MBPS items. There were very weak inter-item correlations between face and movement at 2 months for the period immediately after the needle. This result is problematic for the assumption that all items measure the same construct, that of pain. There were also excessively high correlations between MBPS items at 2 months (face and cry) and 12 months (face and body movement) for the period 1 minute after the needle. Excessive correlations suggest redundancy that inflates internal consistency estimates. Given that MBPS items are not behaviourally coded at fine-grained levels, as in the NFCS, global judgments of pain are more likely to affect ratings on the items. 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 1 and 3 item versions of MBPS. Their convergences between different versions of the scale suggest that a more parsimonious version of each measure (with better data supporting a unidimensional factor of pain across ages and immunization appointment stages) is a solid recommended practice. Moreover, it is important to note that the original 7-item version of the NFCS was not reflective of a unitary construct, lending a serious threat to construct validity. 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. Further, 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, evidenced by very high correlations particularly during the post-needle regulation period. 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 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). Considering the relationships among MBPS items, the associations with NFCS total scores, and using our informed judgment based on 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

The current findings provide important implications for using MBPS and NFCS to assess acute pain in infants. First, the results of this study support the use of the three-item NFCS as an instrument to assess acute pain in preverbal infants, particularly for research purposes, but also perhaps for clinical purposes. With ten or seven facial actions to code, the reliability and coding process can be rigorous and lengthy. The revised version of NFCS (with only three facial actions, confirmed unitary factor structure, solid internal consistency, and convergent validity) 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. 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. Second, 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. Considering all the results together to make recommendations for behavioural assessment of pain related distress in infancy, if time permits the use of NFCS with three facial actions should be used, but if facial activity cannot be coded, assessing cry on the MBPS can be a more efficient way to accurately assess infant pain-related distress.

References

Ahola Kohut, S., & Riddell, R. P. (2009). Does the neonatal facial coding system differentiate between infants experiencing pain-related and non-pain-related distress?. *The Journal of Pain*, *10*(2), 214-220.

Ahola Kohut, S., Pillai Riddell, R., Flora, D. B., & Oster, H. (2012). A longitudinal analysis of the development of infant facial expressions in response to acute pain: Immediate and regulatory expressions. *PAIN®*, *153*(12), 2458-2465.

Biringen, Z. (2008). The Emotional Availability (EA) Scales, 4th Edition. Published by www.emotionalavailability.com.

Bollen, K.A. (1989). Structural equations with latent variables. New York: Wiley.

Campbell, L., Riddell, R. P., Garfield, H., & Greenberg, S. (2013). A cross-sectional examination of the relationships between caregiver proximal soothing and infant pain over the first year of life. *PAIN®*, *154*(6), 813-823.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, *7*(3), 309.

Cohen, L. L., Bernard, R. S., McClelland, C. B., & MacLaren, J. E. (2005). Assessing medical room behavior during infants' painful procedures: The measure of adult and infant soothing and distress (MAISD). *Children's Health care*, *34*(2), 81-94.

Cong, X., McGrath, J. M., Cusson, R. M., & Zhang, D. (2013). Pain assessment and measurement in neonates: an updated review. *Advances in Neonatal Care*, *13*(6), 379-395.

Craig, K. D., Hadjistavropoulos, H. D., Grunau, R. V., & Whitfield, M. F. (1994). A comparison of two measures of facial activity during pain in the newborn child. *Journal of Pediatric Psychology*, *19*(3), 305-318.

Craig, K. D., Whitfield, M. F., Grunau, R. V., Linton, J., & Hadjistavropoulos, H. D. (1993). Pain in the preterm neonate: behavioural and physiological indices. *Pain*, *52*(3), 287-299.

Cunliffe, M., & Roberts, S. A. (2004). Pain management in children. *Current Anaesthesia & Critical Care*, *15*(4), 272-283.

Din, L., Riddell, R. P., & Gordner, S. (2008). Brief report: maternal emotional availability and infant pain-related distress. *Journal of pediatric psychology*, jsn110.

Ekman, P., & Friesen W. (1978). *Investigator's guide to the Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Flora, D.B., LaBrish, C., & Chalmers, R.P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology (Quantitative Psychology and Measurement), 3* (55).

Gélinas, C., Loiselle, C. G., LeMay, S., Ranger, M., Bouchard, E., & McCormack, D. (2008). Theoretical, psychometric, and pragmatic issues in pain measurement. *Pain Management Nursing*, *9*(3), 120-130.

Grunau, R. V., & Craig, K. D. (1987). Pain expression in neonates: facial action and cry. *Pain*, *28*(3), 395-410.

Harrison, D. M. (2013). Naturalistic studies of procedural pain management in infants: Is it ethical to not provide pain management?. *Pain*, *154*(8), 1481-1482.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Johnston, C. C. (1998). Psychometric issues in the measurement of pain. *Progress in pain research and management*, *10*, 5-20.

Johnston, C., Campbell-Yeo, M., Fernandes, A., & Ranger, M. (2013). Neonatal Pain. In *Comprehensive Treatment of Chronic Pain by Medical, Interventional, and Integrative Approaches* (pp. 1015-1029). New York: Springer.

Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Lilley, C. M., Craig, K. D., & Grunau, R. E. (1997). The expression of pain in infants and toddlers: developmental changes in facial action. *Pain*, *72*(1), 161-170.

Lisi, D., Campbell, L., Riddell, R. P., Garfield, H., & Greenberg, S. (2013). Naturalistic parental pain management during vaccinations during the first year of life: Observational norms from the OUCH cohort. *PAIN®*, *154*(8), 1245-1253.

MacCallum, R.C. (2009). Factor analysis. In R.E. Millsap & A. Maydeu-Olivares (Eds), The SAGE Handbook of Quantitative Methods in Psychology (pp. 123-147). Thousand Oaks, CA: SAGE.

Mamedova, K., Pillai Riddell, R., Flora D. (2017). *Parental Report of Self and Child Worry During Acute Pain: A Critical Factor in Determining Parental Pain Judgment.* Manuscript submitted for publication.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … & de Vet, H. C. (2012). COSMIN checklist manual. *Amsterdam: University Medical Center*.

Oberlander, T. F., Grunau, R. E., Whitfield, M. F., Fitzgerald, C., Pitfield, S., & Saul, J. P. (2000). Biobehavioral pain responses in former extremely low birth weight infants at four months' corrected age. *Pediatrics*, *105*(1), e6-e6.

Pillai Riddell, R., Campbell, L., Flora, D. B., Racine, N., Osmun, L. D., Garfield, H., & Greenberg, S. (2011). The relationship between caregiver sensitivity and infant pain behaviors across the first year of life. *Pain*, *152*(12), 2819-2826.

Pillai Riddell, R. R., & Craig, K. D. (2007). Judgments of infant pain: The impact of caregiver identity and infant age. *Journal of pediatric psychology*, *32*(5), 501-511.

Pillai Riddell, R., Fitzgerald, M., Slater, R., Stevens, B., Johnston, C., & Campbell-Yeo, M. (2016). Using only behaviours to assess infant pain: a painful compromise?. *Pain*, *157*(8), 1579-1580.

Pillai Riddell, R., Flora, D. B., Stevens, S., Greenberg, S., & Garfield, H. (2014). The role of infant pain behaviour in predicting parent pain ratings. *Pain Research and Management*, *19*(5), e124-e132.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2),1–36.

Peters, J. W., Koot, H. M., Grunau, R. E., de Boer, J., van Druenen, M. J., Tibboel, D., & Duivenvoorden, H. J. (2003). Neonatal Facial Coding System for assessing postoperative pain in infants: item reduction is valid and feasible. *The Clinical journal of pain*, *19*(6), 353-363.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences*, *2*(1), 13-43.

Sekhon, K. K., Fashler, S. R., Versloot, J., Lee, S., & Craig, K. D. (2017). Children's Behavioral Pain Cues: Implicit Automaticity and Control Dimensions in Observational Measures. *Pain Research and Management*, *2017*.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.

Stevens, B., Franck, L., Gibbins, S., McGrath, P. J., Dupuis, A., & Yamada, J. (2007). Determining the structure of acute pain responses in vulnerable neonates. *CJNR (Canadian Journal of Nursing Research)*, *39*(2), 32-47.

Stevens, B. J., Pillai Riddell, R. R., Oberlander, T. E., & Gibbins, S. (2007). Assessment of pain in neonates and infants. In *Pain in neonates and infants*. Edinburgh: Elsevier Unlimited.

Taddio, A., Hogan, M. E., Moyer, P., Girgis, A., Gerges, S., Wang, L., & Ipp, M. (2011). Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing vaccination injections. *Vaccine*, *29*(7), 1390-1394.

Taddio, A., Nulman, I., Koren, B. S., Stevens, B., & Koren, G. (1995). A revised measure of acute pain in infants. *Journal of pain and symptom management*, *10*(6), 456-463.

Taddio, A., Stevens, B., Craig, K., Rastogi, P., Ben-David, S., Shennan, A., ... & Koren, G. (1997). Efficacy and safety of lidocaine–prilocaine cream for pain during circumcision. *New England Journal of Medicine*, *336*(17), 1197-1201.

Tsze, D. S., von Baeyer, C. L., Bulloch, B., & Dayan, P. S. (2013). Validation of self-report pain scales in children. *Pediatrics*, *132*(4), e971-e979.

Appendix A
Recruitment and data collection procedures

Upon arrival at the pediatrician clinic, caregivers whose infants were undergoing immunizations were offered a study flyer by the clinic receptionist and asked if they would like to learn more about the study. If the caregiver expressed interested in the study, he or she was approached by a research assistant who described the longitudinal nature of the study, explained that data collected would be kept confidential, and informed the participants that they could withdraw from the study at any time. After explaining the study and the consent forms, the caregiver independently reviewed the consent forms and the research assistant answered any questions the caregiver may have had regarding the study's procedures. If the caregiver agreed to participate, he or she signed consent forms and completed a demographic information form with the research assistant. Caregivers and infants were seen by the pediatrician in clinic rooms where the research assistants set up video cameras (see Apparatus below) to capture caregiver sensitivity and infant pain behaviours. Caregiver and infants were videotaped for up to 3 minutes pre-immunization and up to 5 minutes post-immunization. To ensure that the time of the needle was accurately recorded for coding purposes, the research assistant said "now" into the video camera microphone at the moment the infant's skin was punctured by the needle.

Videotaping and completion of demographic questionnaires took place at 2-, 4-, 6- and 12-month immunization appointments. During pilot testing of the study, caregivers indicated that they would like a copy of the immunization videos. As such, caregivers were mailed a copy of their child's immunization video following each appointment. Caregivers were also mailed a postcard reminder and called by research assistants two weeks before their child's upcoming immunization appointment to remind them that a research assistant would be at the appointment to videotape them.

Apparatus

Two Canon HD Video Camcorders – HV20 were used to videotape the caregiver-infant dyads during the immunization appointment. One camera was placed on a tripod and fitted with a wide-angle lens to capture caregiver sensitivity and infant pain behaviours. The second camera was hand-held by a research assistant and recorded a close-up image of the infant's face for subsequent coding of infant facial expression

Appendix B
Neonatal Facial Coding System (NFCS) with seven indicators

Facial Action	**Description**
Brow Bulge	Bulging, creasing and vertical furrows above and between brows occurring as a result of the lowering and drawing together of the eyebrows.
Eye Squeeze	Identified by the squeezing or bulging of the eyelids. Bulging of the fatty pads about the infant's eyes is pronounced.
Nasolabial Furrow	Primarily manifested by the pulling upwards and furrow deepening of the nasolabial furrow (a line or wrinkle that begins adjacent to the nostril wings and runs down and outward beyond the lip corners).
Open Lips	Any separation of the lips.
Vertical Stretch Mouth	Characterized by a tautness at the lip corners (vertical) coupled with a pronounced downward pull of the jaw. Often stretch mouth is seen when an already wide open mouth is opened a fraction further by an extra pull at the jaw.
Horizontal Stretch Mouth	Appears as a distinct horizontal pull at the corners of the mouth.
Taut Tongue	Characterized by a raised, cupped tongue with sharp tensed edges. The first occurrence of taut tongue is usually easy to see, often occurring with a wide open mouth. After this first occurrence, the mouth may close slightly. Taut tongue is still scoreable on the basis of the still-visible tongue edges.

(Grunau & Craig, 1987)

Appendix C

Modified Behaviour Pain Scale (MBPS)

Observed Behaviour	Score	Operational Definition
Facial Expression		
Definite positive expression	0	Smiling
Neutral expression	1	
Slightly negative expression	2	Brow bulge, nasolabial furrow
Definite negative expression	3	Brow bulge, nasolabial furrow, eyes closed tight, open lips, with or without reddened face
Cry		
Laughing or giggling	0	
Not crying	1	
Moaning, quiet vocalizing, gentle or whimpering cry	2	
Full lunged cry or sobbing	3	
Full lunged cry, more than baseline cry	4	To be scored only if infant is crying during baseline
Body Movement		
Usual movements/activity, or resting/relaxed	0	
Partial movement or attempt to avoid pain by withdrawing the limb where puncture is done	2	Squirming, arching, limb tensing/clenching
Agitation with complex movements involving the head, torso, or other limbs, or rigidity	3	Generalized limb and/or body movements, or rigidity

(Taddio, Nulman, Koren, Stevens, & Koren, 1995)

Appendix D

Emotion Availability Scales-Fourth Edition (EAS) caregiver behaviours

EA Adult Sensitivity

#	Subscale	Range	Score
1	Affect	1-7	
2	Clarity of perceptions and appropriate responsiveness	1-7	
3	Awareness of timing	1-3	
4	Flexibility, variety, and creativity in modes of play or interaction	1-3	
5	Acceptance	1-3	
6	Amount of Interaction	1-3	
7	Conflict Situations	1-3	
-	Total (score 7-29)	-	

EA Adult Structuring

#	Subscale	Range	Score
1	Provides appropriate guidance and suggestions	1-7	
2	Success of attempts	1-7	
3	Amount of Structure	1-3	
4	Limit setting, setting boundaries proactively	1-3	
5	Remaining firm in the face of pressure	1-3	
6	Verbal vs. nonverbal structuring	1-3	
7	Peer vs. adult role	1-3	
-	Total (score 7-29)	-	

EA Adult Non-intrusiveness

#	Subscale	Range	Score
1	Follow child's lead	1-7	
2	Non-interruptive ports of entry into interaction	1-7	
3	Commands, directives	1-3	
4	Adult talking	1-3	
5	Didactic teaching	1-3	
6	Physical vs. verbal interferences	1-3	
7	The adult is made to "feel" or "seem" intrusive	1-3	
-	Total (score 7-29)	-	

EA Adult Non-hostility

#	Subscale	Range	Score
1	Adult lacks negativity in face or voice	1-7	
2	Lack of mocking, ridiculing, or other disrespectful statement and/or behavior and general demeanor, whether obvious or subtle	1-7	
3	Lack of threats of separation	1-3	
4	Does not lose cool during low and high	1-3	

	challenge/stress times	
5	Frightening behavior/tendencies	1-3
6	Silence	1-3
7	Themes or play themes hostile	1-3
-	Total (score 7-29)	-

(Biringen, 2008)

Appendix E
Measure of Adult and Infant Soothing and Distress Coding System (MAISD) caregiver behaviours

Adult Behaviour	**Definitions and Examples**
Physical Comfort	Any physical (ie. nonverbal) behavior conducted in an attempt to comfort the child. This may include rubbing, massaging, or patting (on any body part), kissing the child, or a comforting hug. if the adult is simply holding the child so that the procedure may be performed, do not code hug. This has to be an obvious blatant squeeze. Physical comfort can also be coded when a child is being held closely in (e.g. hand pulling head into mom with mom's cheek or chin resting on baby). It is also coded if the child is picked up right after the needle (unless the doctor tells the parents to pick up the baby).
Rocking	If the parent remains in the chair and begins to sway, rock, or bounce the child. When the adult stands up and rocks, sways, or bounces, or when the adult moves around the room while holding the child. The rocking needs to be purposeful. It cannot be coded if caregiver is just walking around the room to get something or adjusting baby's position.
Verbal Reassurance	Reassuring comments (e.g. "it is okay", "we are almost done", "it's alright, baby", "I'm sorry").

(Cohen, Bernard, McLelland, McLaren, MacLaren, 2005)