THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

SCHOOL OF NURSING

INFANT PAIN AND STRESS ASSESSMENT
AN INTEGRATIVE REVIEW

MOLLY K. MCINNIS

Spring 2010

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Nursing and Visual Arts
with honors in Nursing

Reviewed and approved* by the following:

Kim Kopenhaver Haidet
Assistant Professor of Nursing
Thesis Supervisor

Donna M. Fick
Associate Professor of Nursing
Honors Adviser

*Signatures are on file in the Schreyer Honors College.

ABSTRACT

Accurate infant pain and stress assessment is paramount to the appropriate management of pain and stress states. PURPOSE: To analyze recent validation research concerning infant pain tools and to make recommendations to clinicians and researchers regarding the use of these tools. DESIGN AND METHODS: An integrated review of the literature was conducted in October 2009 using the PubMed database with the search terms *measures of stress* and *pain scales*. Results were limited to validation studies published between 2005-2009 using neonatal samples. SUBJECTS: 13 pain instruments within 19 studies were reviewed. Tools were assessed according to study design and psychometrics available. RESULTS: Of the 13 tools reviewed, 7 used strictly behavioral cues and 6 used multidimensional cues. The 19 studies were conducted internationally and were observational or quasi-experimental in design. Full term, preterm, and mixed age samples were used, and the tools were tested for a wide range of psychometrics, including content, construct, and criterion validity, interrater and intrarater reliability, and internal consistency. Several studies conducted nurse reports of feasibility. DISCUSSION: Several pain/stress tools emerged as the most suitable for research and clinical practices, as well as for the purpose of assessing procedural and prolonged pain. Several studies tested for the difference between pain and stress. It is important that future validation studies for infant pain include more accurate reporting of descriptive statistics, tool psychometrics, rater training on the use of the tool, and the gestational age and postnatal age of infant subjects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Kim Kopenhaver Haidet for the expert guidance she has provided me during the process of writing this thesis. Her knowledge about the subject of neonates and her willingness to teach has been imperative to the completion of this work. I am very grateful for her feedback and for her encouragement as I navigated a subject that would have otherwise been overwhelming.

I would also like to thank Dr. Donna Fick for her encouragement and for educating me about the process of thesis writing. I greatly benefited from her guidance and her knowledge of nursing research. I am grateful for her feedback and her support.

Chapter 1: Introduction

In a world of advanced medicine and technology, healthcare providers can determine the accurate vital status of patients by glancing at a monitor. Most vital signs can be taken electronically and can accurately describe patient status in combination with patient assessment. Less advanced is the current determination of pain, a subjective experience of the patient that is often referred to as the fifth vital sign. Certain signs like facial grimacing, increased heart rate, increased blood pressure, and diaphoresis (Jansen, 2008) can alert healthcare providers to the possibility of unreported pain in the adult, but no indicator is as accurate as self-report. Unfortunately, several populations are incapable of standard communication, such as infants, patients with dementia, and those with various communicative disorders.

In the infant population specifically, accurate pain assessment is imperative. Studies suggest that neonatal experiences of pain may be associated with lasting effects, as evidenced by a lower pain threshold at three months of age (Abdulkader, Freer, Garry, Fleetwood-Walk, & McIntosh, 2008b), damage to individual dermatomes (Peters, Schouw, Anand, van Dijk, Duivenvoorden, & Tibboel, 2005), and a lower threshold to stress at school-age (Hohmeister, Demirakca, Zohsel, Flor, & Hermann, 2006). Additionally, pain may have more lasting effects on the preterm population, and compounded stress responses may increase pain experience in this population (Abdulkader et al., 2008b). Several pain instruments have been developed by combining physiological and behavioral indicators of pain, but none are definitive in assessing pain.

In addition to pain, infant stress is also difficult to assess in the non-verbal patient. Fewer instruments have been created to distinguish infant stress states in addition to infant pain, even though infant stress states have also been shown to have a negative effect on infant state (Anand,

1998; Anand, 2007) and basic NICU conditions can induce this stress (Peng, Bachman, Jenkin, Chen, Chang, Chang, & Wang, 2009). In adults, stress has been shown to have an effect on cardiovascular disease, sleep cycles, and memory (McEwen, 1998). Within the infant population, an increased sympathetic arousal state has also been correlated with incidence of Sudden Infant Death Syndrome (SIDS) (Kahn et al., 2003). Fortunately, the application of developmental care for infants has been shown to decrease poor outcomes and stress behaviors in the NICU (Als et al., 2003; Lawhon, Duffy, McAnulty, Gibes-Grossman, & Blickman, 1994; Becker, Grunwald, Moorman, & Stuhr, 1993). However, with such serious complications, knowledge of pain and stress assessment is key.

Previously, Duhn and Medves (2004) organized available infant pain assessment instruments and reviewed their validity and reliability. New tools have been developed since that publication, and the growing literature related to stress assessment instruments was not mentioned in the systematic review (Duhn & Medves). The purpose of this review is to (1) compare infant pain and stress instruments and make recommendations about the reliability, validity, and clinical feasibility of current pain and stress tools in both the research and clinical setting, (2) make recommendations about validation study design to best prove reliability and validity, and (3) discuss the implications for researchers and clinicians regarding the most appropriate cues for assessing pain and stress in preterm infants to allow for appropriate pharmacological and non-pharmacological intervention.

Chapter 2: Review of Literature

Inaccurate pain and stress assessment for the neonatal and infant populations can lead to either insufficient or excessive management of their condition. Knowledge of the infant pain and stress behaviors and the use of a valid and reliable assessment tool are paramount to accurate assessment. In this chapter, the basis of pain and stress behaviors in infants will be described. Additionally, the basic concepts of validity and reliability and the process of creating pain and stress measurement scales with these psychometrics will be outlined.

Introduction to Stress

Previously, literature concerning infant assessment mostly focused on pain as a deleterious condition and ignored stress. However, recent literature has shown that inflicted stress on an infant, often operationally defined as cluster care, can cause much of the same behavioral and physiological cues as pain. Infants take a longer time to return to a baseline status of behaviors after a stressful experience than from a painful experience (Holsti, Grunau, Oberlander, Whitfield, & Weinberg, 2005), and infants who have undergone repeated painful procedures may exist in a continuous state of stress, as indicated by elevated heart rate, and a dampened facial reaction (Grunau, Oberlander, Whitfield, Fitzgerald, Morison, & Saul, 2001). Also, stressful activities performed on infants thirty minutes prior to a painful procedure can lead to more intense pain reactions than pain procedures performed after a period of undisturbed rest (Holsti, Grunau, Oberlander, & Osiovich, 2008). To understand stress in an infant, the concept of general stress should be explained.

As first defined by Hans Selye in 1936, stress is a three-step reaction to a stimuli, known as the General Adaptation Syndrome (GAS). During the first step, called the *alarm* stage, the body has a general response including the *fight or flight* response, in which the body enables all

processes needed for possibly intense physical activity. The sympathetic nervous system is enabled, which causes increased heart rate, increase blood pressure, diaphoresis, decreased bowel motility, and other excitatory effect. The inhibitory parasympathetic nervous system is suppressed. The second stage is called the resistance phase, in which the body continually fights for homeostasis against stressors. In order to maintain the energy to fight the stressor, body states like blood sugar and blood pressure may increase. If the body cannot maintain the resistance phase, it enters the exhaustion stage, where homeostasis may no longer be possible (Selye, 1936). In application to the newborn, the resistance stage may be short, especially in the preterm infant who does not have the energy reserve to maintain high blood sugar without using precious brown fat reserves and converting to anaerobic metabolism. Maintenance of a resistance stage and anaerobic metabolism can start a dangerous cycle toward acidosis and even death (Cornblath & Ichord, 2000).

As has been discussed in recent literature, stress is not an exclusively negative condition. Bruce McEwen, an expert on the physiological state of stress, compares Selye's GAS to the concepts of allostasis and allostatic load. Allostasis is the maintenance of the most integral and least compensating body markers like pH, temperature, and oxygen levels. These values have to remain rather stable to support bodily functions. They are maintained through the variability of more adjustable systems, or allostatic systems, like blood pressure and cortisol levels. These allostatic systems may change to ensure the homeostasis of the entire system. The introduction of an allostatic state is comparable to the resistance stage of Selye's GAS; allostasis can persist until factors like blood pressure can no longer be maintained, and allostatic overload, or the exhaustion phase of the GAS ensues (McEwen, 2005). Because some body systems are

immature in comparison to an adult, an infant's allostasis may not be maintained with especially intense stressors.

Selye's and McEwen's ideas of stress and the maintenance of homeostasis can be related to Heidelise Als' research and her concept of developmental care (Als, 1982; Als et al., 1994). According to Als, preterm infants use (1) autonomic, (2) motor, (3) state-specific, and (4) attention as regulatory strategies to maintain the allostasis introduced by McEwen. These strategies begin early in fetal life and allow the infant to react and adapt to stressors during development. The uterine environment allows for the protected maturation of these five processes as the pregnancy continues to term. However, infants who are born prematurely are forced to continue development in a harsher environment than expected. When infants have underdeveloped regulatory processes they are unable to counter even the slightest stressors, and can be launched into allostatic overload. With Als' concept of developmental care, an infant's stress-regulation is assessed in relation to the four regulatory processes previously mentioned, and care is organized according to that assessment to create the least taxing physical environment possible. As shown in a 1994 study of 38 preterm infants, the use of developmental care significantly decreases negative outcomes, including the number of days an infant requires supplemental oxygen and tube feeding, the length of hospitalization, the amount of hospital charges, and the incidence of bronchopulmonary dysplasia and intraventricular hemorrhage (Als et al., 1994). In a study of 92 preterm infants, Als and colleagues found that developmental care also led to shorter stays in intensive care, fewer cases of necrotizing enterocolitis, lower ages at discharge, better regulatory functioning, improved familial interaction, and better growth as judged by length, weight, and head circumference (2003). In addition to developmental care,

healthcare providers use other pharmacological and non-pharmacological interventions to reduce stress.

Non-pharmacological interventions for stress include several types of comfort measures: skin to skin contact, often referred to as *kangaroo care* (Ludington-Hoe, Hosseini, & Torowicz, 2005), non-nutritive sucking, breast feeding (Shah, Aliwalas, & Shah, 2007), and massage (Jain, Kumar, & McMillan, 2006). Because these forms of intervention are not harmful to the child, they are also used for infants in pain. Pharmacological interventions for infant stress include sedatives like midazolam, a benzodiazepine, and ketamine, an N-methyl-D-aspartic acid (NMDA) antagonist. These sedatives are more often given to infants who are mechanically ventilated than those who are not (Anand, 2007).

Introduction to Pain

In the past, some clinicians believed that neonatal nociception systems were not mature enough to transmit messages of pain. K.J.S. Anand, an expert in the field of neonatal pain, has written many reports explaining the evidence that the neurological systems involved in pain nociception are mature enough to transmit pain signals in both preterm and term neonates (Anand, 1998; Anand, 2007; Anand, Sippell, & Aynsley-Green, 1987). Neonatal experiences of pain can have long-term negative effects. In a 2008 study, Abdulkader and colleagues found that infants who experienced a repeated painful event, operationally defined as a heel stick, had a lower threshold for painful stimuli in the same dermatome three to five weeks later after the stick (Abdulkader et al., 2008b). Other studies have demonstrated similar results when assessing pain in dermatomes in which pain has previously been inflicted (Oberlander, Gruanu, Whitfield, Fitzgerald, Pitfield, & Saul, 2000; Peters et al., 2005). Effects of pain may last years, as shown in a study by Hohmeister (2006) in which the reaction to prolonged painful stimuli was

examined in preterm and full term school-aged children and the results were organized according to which children had been exposed to multiple painful procedures during infancy. According to Hohmeister, children ages nine through fourteen years who had experienced painful procedures as infants had lower thresholds for heat pain (Hohmeister et al., 2006). As recent literature has proven, pain has lasting neurological effects on infants. To decrease the effects of pain in this population, several form of pharmacological analgesia are available for administration.

Previously, even when infant pain was addressed as a possibility, the experience of pain was thought to be less damaging to the developing child than that the use of analgesia. Many clinicians believed infants did not feel pain, and some continue to question its existence (Derbyshire, 1999). However, the majority of researchers maintain that infant pain is real and have investigated different forms of pharmacological analgesia to reduce infant pain. Some of the most common forms of analgesia include morphine, fentanyl, ketamine, midazolam, and sucrose. Morphine is the most popular choice and in a study of long term administration in infants it had no deleterious long term effects (MacGregor, Evans, Sugden, Guassen, & Levene, 1998). Morphine is efficient, but can cause hypotension, respiratory depression, and can slow bowel motility (Anand, 2007). Morphine also may not be effective in ventilated preterm infants for procedural pain (Carbajal, Lenclen, Jugie, Paupe, Barton, & Anand, 2007). Fentanyl, another option for opioid analgesia, is about 20-30 times more potent than morphine and has a quick, three minute onset time. However, there is stronger evidence for withdrawal (Anand, 2007). Ketamine, an NMDA antagonist, and midazolam, a benzodiazepine, are also administered before painful procedures. They are considered sedatives, not analgesics, and as such do not provide relief of pain, but may be important adjuvant therapy.

In addition to the use of opioids and sedatives, administering sucrose before procedural

pain has been shown to decrease pain response in infants (McCullough, Halton, Mowbray, & MacFarlane, 2008). One study in particular found that sucrose relieved procedural pain, although a stress response was still seen by increased heart rate, oxygen consumption, and energy needs (Bauer, Ketteler, Hellwig, Laurenz, & Versmold, 2004). These forms of pain relief can be administered when infants undergo procedures that have been found to be painful, or when they exhibit behavioral and physiological pain responses. For all pharmacological and non-pharmacological interventions, accurate pain assessment is key in order to evaluate the efficacy of the intervention. This assessment requires the use of accurate and reliable pain assessment tools.

Commonly Used Items in Pain and Stress Assessment Tools

Pain and stress assessment tools combine cues that are associated with infants in pain and who have stress, such as facial expressions, cry quality, body movement, body tone, state/sleep, vital signs, and gestational age.

Facial expression.

The idea that infant facial expressions could be catalogued to assess infant pain and general mood was first introduced by Izard and colleagues through the Maximally Discriminative Facial Movement Coding System (Izard, 1979), which includes twenty-seven different facial movements that could be correlated with a noxious event. This system coded for a range of emotions in addition to pain, including interest, joy, and surprise. Since the creation of that tool, certain facial expressions have repeatedly been tested with pain in neonates.

More recently, the Neonatal Facial Coding System (NFCS) has detailed facial characteristics specific to pain (Grunau & Craig, 1987). These characteristics include (1) brow bulge, (2) eye squeeze, (3) deepened nasolabial furrow, (4) open lips, (5) vertical stretched

mouth, (6) horizontal stretched mouth, (7) lip purse, (8) taut tongue, (9) chin quiver, and (10) tongue protrusion. Some of these facial cues have proven more salient in pain assessment; five have been consistently associated with infants in pain: brow bulge, eye squeeze, nasolabial furrow, open lips, and taut tongue (Grunau & Craig, 1987). In 2003, Peters and colleagues showed that three of the NFCS cues could assess pain as reliably as the expanded tool. Nasolabial furrow, brow bulge, and eye squeeze were found to be as predictive of pain as the ten-item tool (Peters, Koot, Grunau, de Boer, van Druenen, & Tibboel, 2003). Another study found that infants with varying neurological deficits could still be assessed consistently with brow bulge, facial grimace, eye squeeze (Stevens et al., 2006).

Many of the elements of the NFCS have been tested with infants undergoing stressful procedures as well as painful procedures. Several studies have found that the presence of brow bulge and nasolabial furrow increase significantly with non-pain events like cluster care (Grunau, Johnston, & Craig, 1990; Holsti, Grunau, Oberlander, Whitfield, & Weinberg, 2005). While it seems the validity of the NFCS as a pain tool would be compromised by the positive identification of stress, these studies have shown that when taking into consideration the constellation of items, painful situations elicit higher overall scores from infants than do stressful situations (Grunau, Johnston, & Craig, 1990; Pereira et al., 1999). However, no single item on the NFCS can consistently define stress independent of pain or vice versa.

Additional pain and stress assessment instruments incorporate elements of the NFCS. Researchers Stevens, Johnston, Petryshen and Taddio used nasolabial furrow, brow bulge, and eye squeeze from the NFCS when creating the Premature Infant Pain Profile (PIPP) (Stevens, Johnston, Petryshen, & Taddio, 1996), while Holsti and Grunau used brow bulge, eye squeeze, nasolabial furrow, horizontal stretched mouth, and taut tongue in the Behavioral Indicators of

Infant Pain scale (BIIP) (Holsti & Grunau, 2007b). These tools were both created with preterm infant samples, and their choice of facial cues reflects that population. Preterm infants typically have less facial reactivity to pain than full term infants (Gibbins et al., 2008), but the facial characteristics that they do express typically include more upper facial expression than lower facial expression due to their progress in cephalocaudal development (Haidet et al, 2009). Additionally, the BIIP may use horizontal stretch mouth because it is more common than vertical stretch mouth in preterm infants, perhaps due to the fact that horizontal movement requires less advanced muscular effort (Johnston, Stevens, Craig & Grunau, 1993).

Grimace is another facial cue used for infant pain and stress assessment. In the *C*rying, *R*equires oxygen, *I*ncreased vital signs, *E*xpression, and *S*leepless (CRIES) assessment scale (Krechel & Bildner, 1995) the term grimace is described as "lowered brow, the eyes squeezed shut, a deepening of the nasolabial furrow, and open lips and mouth" which effectively describes the assembled cues of the NFCS (Pasero, 2002).

Cry.

Another item often used to assess for pain and stress in infants and neonates is *cry*. Cry is not a consistent factor in pain and stress assessment because it can occur in response to or independently of pain. Also, some infants are not able to cry when they feel pain or stress. According to Pereira and colleagues, only 50% of infants cry in response to a painful event (Pereira, Guinsburg, de Almeida, Monteiro, dos Santos, & Kopelman, 1999). Also, preterm infants are less likely to cry in response to pain in comparison to full term infants, and their cry is significantly higher in pitch (Manfredi, Bocchi, Orlandi, Spaccaterra, & Donzelli, 2009). Some infants may not even be able to cry in painful situations due to an increased stress response

(Sredl, 2003). Despite its lack of consistency, one can use the presence of cry as a cue to assess pain when considered alongside other possible pain cues.

Cry is also inconsistent as a pain cue due to the fact that cry cannot be audibly interpreted in neonates who are intubated. Intubated infants that are completely mechanically ventilated are not capable of producing sound through their vocal chords, but they are still able to produce a *cry face*, which includes the same brow bulge, nasolabial furrow, and eye squeeze of an infant who can audibly cry. As a result, tools which include cry as an item may interpret the facial indicators of cry in the case of ventilated infants. Both the modified Face, Legs, Activity, Cry, and Consolability scale (FLACC) (Johansson & Kokinsky, 2009; Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997) and the COMFORT behavioral scale (van Dijk, de Boer, Koot, Tibboel, Passchier, & Duivenvoor-den, 2000) use the cue *cry-face*.

Several researchers have studied cry characteristics, including the cry pitch, the length of a cry, and the way it oscillates, or whether it is a *siren cry* (Bellieni, Sisto, Cordelli, & Buonocore, 2004; Stevens, Johnston, & Horton, 1994). Analyzing cry characteristics is the basis of the ABC (*A*cuteness of the first cry, *B*urst rhythmicity, and *C*onstancy of cry) scale for infant pain assessment (Bellieni, Bagnoli, Sisto, Neri, Cordelli, & Buonocore, 2005). A recent study researching cry characteristics found that high NFCS scores correlated with the highest pitch and volume cries in preterm infants, although these changes were analyzed electronically, as is the case with most research regarding cry analysis (Luhr, Zeskind, Ofenstein, Cepeda, Warrier, & Aranda, 2007). Identifying infant cry types by ear have become a popular way to recognize whether a child is expressing hunger, stress, or digestive activity (Dunstan, 2009), but these ideas have not been scientifically tested.

Motor signs.

Body movement and body tone are two common motor cue types in pain and stress assessment tools. Als has reported that infants self regulate in response to stress through general flexion of muscles (Als, 1982; Als et al., 2003). Other body movements have also been related to stress; Grunau, Holsti, Whitfield, and Ling found finger splay and hyperextension to be significantly associated with stress (1999). Again, Holsti, Grunau, Oberlander, Whitfield and Weinberg also found an increase in grasping and finger splay in response to stress (2005). Several studies have found leg flexion to be a pain response, but these results may be due to the nature of the pain event, usually defined as trauma to the heel (Fitzgerald, Millard, & MacIntosh, 1988; Abdulkader, Freer, Fleetwood-Walker, & McIntosh, 2008a)

While body movement, specifically flexion, is often considered a stress response (Als, 1982), some studies have noted leg flexion in response to pain. Infants respond to local pain through movement; they have a natural withdrawal reflex in an affected extremity in response to pain or stimulation of that extremity (Fitzgerald et al., 1988). In 2008, Abdulkader and colleagues tested infant motor responses to non-pain heel stimulation and heel stick. Light stimulation led to flexion withdrawal of the heel and a predictable sequence of bodily movements including eye opening, head rotation, chin lift, and gross body movements. Both preterm and full term infants had the same reaction sequence, but preterm infants were more sensitive and responded to much softer stimuli, as did all patients who had previously received heel sticks (Abdulkader et al., 2008a).

State/sleep.

Infant state is often included in assessment tools because it can affect the intensity of an infant's reaction to pain or stress. Common behavioral states include *cry/agitation, active*

awake, *quiet awake*, *active asleep*, and *quiet asleep* (Brazelton, 1973). Infants in deep sleep at the time of a procedural pain event mount a less dramatic facial pain response than infants who are alert and awake before the event (Stevens, Johnston, & Horton, 1994); however the physiological consequences may be greater. Thus, the authors of the Premature Infant Pain Profile (PIPP) account for the unexpressed pain and stress of infants in deep sleep by assigning points on the scale for that state. In most assessment scales, an increase in score indicates increased pain; In the PIPP, an infant receives one point for *quiet/awake*, two points for *active/sleep*, and three points for *quiet/sleep* states (Stevens, Johnston, Petryshen, & Taddio, 1996) because a sleeping infant may not express pain as quickly or vigorously as an alert infant. Authors of other pain and stress tools argue that one cannot assume that an infant in deep sleep is not feeling pain or stress. Holsti and Grunau, creators of the Behavioral Indicators of Infant Pain (BIIP) scale, assign zero points to infants in deep sleep, active sleep, drowsy, and quiet awake states, one point for the active awake state and two points for crying (Holsti & Grunau, 2007b). The differences between the BIIP and PIPP authors' tactics for sleep scoring may be a result of differences in the characteristics of research populations; The creators of the BIIP have dealt more with term infants while the creators of the PIIP have predominantly studied preterm infants. Other tools which describe infant sleep or state as a cue and assign additional points to a more agitated infant include the Neonatal Infant Pain Scale (NIPS) (Lawrence et al., 1993), the Pain Assessment Tool (PAT) (Hodgkinson, Bear, Thorn, & Van Blaricum, 1994), the COMFORT scale (Ambuel, Hamlett, Marx & Blumer, 1992), the Crying, Requires oxygen, Increased vital signs, Expression, and Sleeplessness scale (CRIES) (Krechel & Bildner, 1995), the Échelle Douleur Inconfort Nouveau-né (EDIN) (Debillon, Zupan, Ravault, Magny, & Dehan, 2001), the Neonatal Pain, Agitation, and Sedation Scale (N-PASS) (Hummel, Puchalski, Creech,

& Weiss, 2008), and the Face, Legs, Activity, Cry, and Consolability scale (FLACC) (Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997).

Physiological indicators.

Typically, infants are very physiologically reactive to stress and pain. In response to a heel stick, infant heart rate and blood pressure generally increase, while oxygen saturation and heart rate variability decrease (Franck & Miaskowski, 1997). Physiological markers are less specific to pain than the previously mentioned behavioral cues; tactile handling can increase heart rate and blood pressure when no facial reaction is seen (Ranger, Johnston & Anand, 2007). Physiological markers may indicate stress as much as pain. Also, among preterm infants, Lucas-Thompson and colleagues illustrated that extremely premature infants [28-31 weeks gestational age at birth (GAB)] are more physiologically reactive to painful stimuli than infants born at 32-34 weeks GA when compared immediately after birth and several weeks later (Lucas-Thompson, Townsend, Gunnar, Georgieff , Guiang, Ciffuentes, Lussky, & Poggi Davis, 2008).

Other changes in bodily processes are less readily measurable but are correlated with pain and stress in infants. These markers include cortisol and catecholamine production, vagal tone index, palmar sweating, and skin blood flow (Franck & Miaskowski, 1997).

Pain and stress assessment tools can contain many different combinations of the behavioral and physiological cues described, and may use subjective data such as *nurse opinion*, or objective data such as *gestational age (GA)*. Examples of the cues used in the pain assessment tools included in this review are summarized in Table 1.

Table 1

Types of Pain Cues Present in Each Reviewed Scale

Cues	ABC SCALE	BIIP scale	NFCS	NIPS	PIPP	PAT	COMFORT	COMFORT-b	CRIES	CHIPPS	EDIN	N-PASS	FLACC
Cry	x			x	x	x	x	x	x	x		x	x
Facial Expression		x	x	x	x	x	x	x	x	x	x	x	x
Behavioral State		x		x	x	x	x	x	x		x	x	
Body Movement		x		x			x	x		x	x		x
Body Tone						x	x	x		x		x	
Vital signs				x	x	x	x		x			x	
Gestational Age					x							x	
Consolability											x		x
Nurse perception						x							

The Observable Difference between Pain and Stress

As has been previously mentioned, many healthcare providers once cared for neonates with the assumption that their nervous systems were not developed enough to feel pain (Derbyshire, 1999). At this point in time, that assumption seems preposterous, and clinicians make every effort to treat infant pain. In the same way, the medical community is slowly realizing the incredible toll that compounded stress can have on an infant's body. Whereas a diaper change was once viewed as a harmless procedure, more and more clinicians enforce cluster care and quieter environments in an effort to decrease infant stress (Harmon & McManus, 2008). Some hospitals have enforced Als' developmental care techniques to create individualized care for each infant based on his or her level of functional maturity; the Newborn Individualized Developmental Care and Assessment Program (NIDCAP) promotes limiting

stressors to the individual infants capabilities and has been shown to improve outcomes in critically ill neonates (Als, 1982, Als et al., 1994, Als et al., 2003). As stress has become a more prominent issue for infant care, literature has developed based on assessing stress and distinguishing it from pain.

Many studies have been conducted related to distinguishing stress responses from pain responses in neonates. Physiologically, non-pain stressors can result in significantly different responses from painful stressors. Infants who are stressed may stay in a state of increased sympathetic arousal longer than infants in pain; studies show infants in the recovery stage of a stressful event may maintain tachycardia for a longer period of time than infants recovering from pain (Holsti et al., 2008; Pereira et al., 2008). This difference illustrates the significance of infant stress, but may not be practically useful for definitively differentiating pain from stress.

The role of body movements for distinguishing infant pain and stress has also been explored. In a 2007 review of literature, Holsti and Grunau outlined infant body movements associated with infant stress and pain. None of the cues studied were found to be definitive for pain and stress on their own; however, several were found to increase significantly with pain or stress. General flexion and extension of arms and legs, hand on face, sitting on air, salute, airplane, finger splaying, and fisting were all found to be associated with pain and stress. These movements are consistent with Als' theory that infants use certain motor strategies to react to stressors. Both finger splay and fisting were seen more often in the preterm infant less than 30 weeks gestational age. Most of the cues were seen in both painful and stressful scenarios. Fisting was the only cue which appeared to be more associated with stress than pain, but the authors determined that this cue alone cannot define stress.

In pain and stress assessment, prematurity can also have an impact on the validity of pain measurement. Prematurity can affect infant's reactions to noxious events due to their increased exposure to painful events, increased sensitivity, and premature behavioral reactivity. As previously stated, infants with previous pain experiences are more sensitive to tactile stimulation (Abdulkader et al., 2008). Infants admitted to a NICU due to prematurity are more likely to undergo more painful procedures than a healthy, term infant. In a NICU in the United Kingdom, over 3000 procedures were performed on a group of 54 preterm infants that were consecutively admitted to the unit, and within that preterm sample, significantly more events were performed on the lowest GA infants (Barker & Rutter, 1995). Due to the increased amount of painful events that a preterm infant must undergo, increased sensitivity to tactical stimulation in this population is an unfortunate reality, and one cannot conclusively know that stimuli like cluster care is not painful to the premature infant (Grunau, Johnston, & Craig, 1990). In addition to increased sensitivity, this population manifests pain differently than term infants. Infants may express different facial reactions to pain, such as increased horizontal stretch mouth (Johnston, Stevens, Craig & Grunau, 1993) and may have increased physiological reactivity to stress (Lucas-Thompson et al., 2008).

Based on the current literature, the differences between pain cues and stress cues are not definitive. The ambiguity of the cues may be a result of the inherent ambiguity of the stress and pain! Stress and pain are not mutually exclusive conditions and their possible overlap makes assessment difficult. Further research is needed to establish cues which can better delineate stress and pain, especially in the preterm infant

Tool Assessment: Validity and Reliability

When evaluating pain/stress assessment tools, a basic knowledge of the concepts of validity and reliability is necessary. To determine whether a tool has acceptable validity and reliability, it is important to know the basic forms of these statistical measures and how a tool can be tested for them. Within the theme of validity, a tool can have content, construct, and criterion validity. In terms of reliability, tools are often tested for interrater reliability, intrarater reliability, and internal consistency.

Validity.

An instrument is valid if it measures the idea or construct for which it was created (DeVon et al., 2007). A construct is an abstract concept or disease process that cannot be definitively diagnosed by objective means (Streiner & Norman, 1994, p 151). In relation to the construct of infant pain, a tool is valid if it can identify when an infant is in pain. Because it is impossible to know exactly when an infant is in pain due to lack of self report, a tool must bring together symptoms, behaviors, and cues that are present during an event that is known to cause pain. When using a valid pain tool, clinicians can identify when a patient exhibits the same cues that were seen in a sample population during a painful event. Clinicians can use the tool to decide whether the patient is in pain. Many of these tools have an established numerical score associated with pain (Jonsdottir & Kristjansdottir, 2005; Stevens, Johnston, Franck, Petryson, Jack, & Foster, 1999). Validation studies with controlled pain events and appropriate samples are necessary to decide what that definitive pain score is. The most commonly discussed types of validity are content, construct, and criterion validity.

Content validity.

Content validity is the easiest type of validation to establish because it is the most subjective, but it is also the least definitive. Content validity is the extent to which a tool represents the fullest spectrum of the construct. It is most often formed during the development of a tool. Methods that strengthen content validity include a review of pain literature to identify many possible cues for inclusion, expert opinion on the tool from at least five different professionals established in the field, and factor analysis to remove all cues which are redundant (DeVon et al., 2007; McDowell & Newell, 1996). A tool also has greater content validity if it is multidimensional; a pain tool which assesses both behavioral and physiological cues many have better content validity than one that only assesses behavior (Streiner and Norman, 1994, p 147). The simplest form of content validity is called face validity, in which experts and/or non-experts can assess the tool superficially based on "grammar, syntax, organization, appropriateness, and confirmation that it appears to flow logically" (DeVon et al., 2007).

Construct validity.

Construct validity is the measure of how well a tool measures the idea or *construct* on which it is based (DeVon et al., 2007). In relation to infant pain, construct validity is measured by how well the tool measures pain. This is often tested in studies by *extreme groups*. Extreme groups are populations that definitively either have the proposed construct or do not (Steiner & Norman, 1994). Many researchers developing infant pain tools will study an intervention group during a necessary, routine painful procedure, such as a blood draw from the heel. Tool developers who have a proposed set of cues that they believe best reflect pain will test their tool on the infant before, during, and sometimes after the painful event. Because an infant is assumed to be in pain during a blood draw, the score of the infant's pain using the proposed tool

during the blood draw should be different than the infant's pre-pain assessment. The infants are usually assumed to be pain-free before the heel stick, so their pre-assessment score is their control, or the opposite *extreme group*. Two additional concepts related to construct validity are convergent and discriminant validity.

Convergent and discriminant validity are very similar to the *extreme groups* previously mentioned within construct validity. If a pain tool has convergent validity, it will correlate with other cues associated with the construct. These other cues could be a pain event, lab test, another tool, or any other test or outcome that could be associated with infant pain (Steiner & Norman, 1994). Discriminant validity is the opposite of convergent validity. If a pain tool has discriminant validity, the tool should not correlate with cues that are not associated with infant pain (Steiner & Norman, 1994). McDowell and Newell refer to convergent and discriminant validity as sensitivity and specificity, respectively (McDowell & Newell, 1996). Sensitivity is described as the ability of the tool to diagnose those with a construct as positive for that construct (true positives), and specificity is described as the ability of the tool to define those without the construct (true negatives).

Criterion validity.

Another form of validity is based on a tool's ability to measure a construct in comparison to another tool that is considered a *gold-standard* for the construct. This type of validity is called criterion validity. To measure criterion validity, pain researchers will often test the construct validity of a tool alongside a previously validated tool. Of course, if another tool is already considered a gold-standard for the construct, researchers must explain why the new tested tool is at all necessary. Common reasons include developing a tool that is more cost efficient, time efficient, easier to teach, and more reliable (McDowell, & Newell, 1996). Sometimes pain

researchers have problems with criterion validity based on the types of tools being compared. Although all infant pain tools test the same construct of pain, some employ methods which are less comparable to each other. For example, tools which rely on behavioral cues to assess pain often garner slightly different results than those that use physiological cues and are not immediately comparable (Holsti & Grunau, 2007a). Also, there is often not a tool that can be considered a gold standard for certain constructs, such as for tools measuring both stress and pain in infants. To validate these tools, expert opinion is often the only option for criterion comparison (Wielenga, De Vos, de Leeuw, & De Haan, 2004). Criterion validity can also be divided into concurrent and predictive validity.

Within criterion validity, concurrent validity represents a comparison of a proposed pain tool with a standard, or criterion, at the same moment in time (DeVon et al., 2007). For instance, pain tools assessed alongside one another indicate concurrent validity because their results are available at the same time. Concurrent validity is the more common form of criterion validity in infant pain assessment validation. Predictive validity represents the measurement of a construct that cannot be assessed until a later date (McDowell & Newell, 1996). For example, tools which assess patient outcomes cannot be evaluated for their validity until the patient is assessed at a future time point.

Reliability.

While a tool's validity is its ability to test for the presence of a certain construct, its reliability reflects how consistently it can do so, or the reproducibility of response under the same conditions. The reliability of an instrument is determined by statistically evaluating the agreement between tool items or intraclass correlation. This is most commonly evaluated through Cronbach's alpha for each tool item. The closer the score approaches 1.0, the higher the

degree of correlation. Reliability can be tested in terms of a tool's items and raters by testing for internal consistency and rater reliability, respectively.

Internal consistency is the measure of how much the individual items of a tool correlate with each other (Streiner & Norman, 1994). Some tools contain items that, while they measure the same construct, provide divergent data. For example, some tools have both physiological and behavioral cues. These two types of cues may not entirely correlate, which could affect the internal consistency of the tool. However, as mentioned previously, a wide range of cues increases content validity which, according to Streiner and Norman, is more important than internal consistency (Streiner & Norman, 1994).

Infant pain researchers are often concerned about the reliability of raters, as misinformed raters can cause significant systematic error. Pain tools may be used by a wide range of people with different educational and cultural backgrounds and varying levels of experience. In a reliable tool, rater attributes should not significantly affect the scores. To describe the reliability of the tool between raters, researchers measure interrater reliability and intrarater reliability. Interrater reliability is the level of agreement between two raters' assessment of the same event (Tinsley & Weiss, 1975). It can be proven through several different statistical tests, and the test that is employed depends on the attributes of the sample. Common tests for interrater reliability are intraclass correlation, Pearson's *r*, Cohen's kappa, and Cronbach's alpha (Abdei, 1996). Intrarater, or within rater, reliability assesses the same rater twice; it is the agreement between one rater's scores at two different times. To measure intrarater reliability, the rater often assesses an infant at bedside or by video, and reassesses that same event by video at a later date (DeVon et al., 2007). Some researchers also test intrarater reliability of video coders by replaying an assessment twice at random during a lengthy coding session. The reliability of the rater during

one session evaluates *rater fatigue*, which is the possible decreased reliability of a rater during a single assessment period.

Descriptions of the tools assessed within this review and their psychometrics are presented in Tables 2-5.

Pain and Stress Assessment Tools

Behavioral and physiological pain cues are the basis for most pain and stress assessment tools. In assessing pain and stress, many studies investigate the use of physiological parameters like heart rate, heart rate variability, oxygen saturation (Haidet, 2005; Stevens et al., 2008), skin conductance (Hellerud & Storm, 2002), salivary cortisol (Herrington, Olomu, & Gellar, 2004), and behavioral indicators like body movements, facial actions, and infant sleep state (Haidet, 2005; Holsti, Grunau, Oberlander, & Whitfield, 2004). Most behavioral indicators can be summarized by NIDCAP, the Newborn Individualized Developmental Care and Assessment Program that includes 61 bodily movements in its assessment (Pressler, Hepworth, Helm, & Wells, 2001). Thirty of these cues, like flexing arms and legs, yawning, and tremors, have been associated with infant pain and stress (Holsti, Grunau, Oberlander, Whitfield, & Weinberg, 2005). All of these behavioral and physiological assessment cues are not definitive enough to imply pain, but a combination of these cues, which is the basis of a pain or stress assessment tool, can lead to better correlation with pain and stress.

There are over thirty pain assessment tools for infants, which differ from the kind of pain assessed to the characteristics of the infant being assessed. Pain can be divided into acute pain, which is a defense mechanism that alerts the body to damage, and chronic pain, which has no purpose for survival. Most pain assessment tools are based on acute pain, which

Table 2

Description of Unidimensional Scales Reviewed

Scale	Creators	Basis	Behavioral Cues	Scoring system
ABC	(Bellieni et al., 2005)	A behavioral scale created to assess pain by cry quality based previous on research (Bellieni et al., 2004)	***A***cuteness of the first cry, ***B***urst rhythmicity and temporal ***C***onstancy of cry intensity	Score 0-6, 2 pts for each cue
BIIP	(Holsti & Grunau, 2007b)	A behavioral tool created for preterm infants using certain NFCS cues (Grunau & Craig, 1987), body movement, and infant state. Cues chosen through ROL and studies of preterm infant pain reaction.	Sleep/wake state, brow bulge, eye squeeze, nasolabial furrow, horizontal mouth, taut tongue, finger splay, fisting	Score 0-9, 1 point per cue, crying/agitation can receive 2 pts.
CHIPPS	(Büttner & Finke, 2000)	A behavioral post-operative pain scale comprised of five cues and created for clinical utility. Büttner & Finke validated it for newborns, infants, and young children.	Crying, facial expression, posture of the trunk, posture of the legs, and motor restlessness.	Score 0-10, 0-2 points for each cue.
COMFORT-B	COMFORT adapted to COMFORT–B by van Dijk et al., 2000)	Adapted from the COMFORT scale, it is a unidimensional tool created to measure distress in PICU patients. A version including only behavioral cues was created in response to divergent behavioral and physiological data results in the original scale. Both ventilated and non-ventilated patients can be assessed with this scale.	Alertness, calmness/agitation, respiratory response (ventilated children) or cry (non-ventilated children), physical movement, muscle tone, facial tension	Score 6-30, 1-5 points for each cue.
EDIN	(Debillon at al., 2001)	A five cue behavioral scale developed for prolonged pain in premature infants. In validation studies, it has been used at the end of an eight hour observation period.	facial expression, body movement, quality of sleep, quality of contact w nurses/sociability, and consolability	Score 0-15, 0-3 points for each cue
FLACC	(Merkel et al., 1997)	A behavioral tool created to assess postoperative pain in preverbal and non-verbal children.	Face, legs, activity, cry, and consolability	Score 0-10, 0-2 points for each cue.
NFCS	(Grunau & Craig, 1987)	A behavioral tool that uses facial actions to assess pain in infants. It is based on the Facial Action Coding System used in adults. It has been validated for postoperative and procedural pain.	10 facial actions: brow bulge, eye squeeze, deepened nasolabial furrow, open lips, vertical stretched mouth, horizontal mouth, lip purse, taut tongue, chin quiver, tongue protrusion. Tongue protrusion is not an indicator of pain in FT infants.	Score 0-9 in FT infants, 0-10 in PT infants, 0-1 points for each cue.

Table 3

Description of Multidimensional Scales Reviewed

Scale	Creators	Basis	Behavioral Cues	Physiological Cues	Scoring system
COMFORT	(Ambuel, Hamlett, Marx, & Blumer, 1992)	COMFORT was developed from the feedback of 20 experienced CCU nurses as a non-obtrusive, multidimensional measure of distress in unconscious and intubated pediatric patients.	alertness, calmness/ agitation, physical movement, muscle tone, facial tension	Respiratory response MAP heart rate	All cues are rated 1-5, with a possible score of 5-40. Sedation earns a lower score, pain a higher score.
CRIES	(Krechel & Bildner, 1995)	CRIES is a five cue neonatal postoperative pain scale created for clinical feasibility.	Crying, expression, and sleeplessness	requires O_2 (SaO_2 below) and "Increased vital signs" (BP and HR)	All cues are rates 0-2, with a possible score of 0-10.
N-PASS	(Hummel, Puchalski, Creech, & Weiss, 2008)	The N-PASS is a multidimensional pain assessment tool created for acute prolonged pain, but also intended for procedural and chronic pain. Negative scores indicate sedation and positive scores indicate pain.	Crying/irritability, behavior/state, facial express-ion, extremities/tone.	Vital signs. Additionally, points are added for GA. 1 point is added for infants <30 weeks GA.	-2 to +2 points for each cue, with a possible score of -10 to +10.
NIPS	(Lawrence et al., 1993)	This procedural pain scale was created to be simple to administer on PT and FT infants with six cues and one non-invasive physiological measurement. It was adapted from the CHEOPS SCALE, which is for older children.	Facial expression, arms (relaxed/ contracted), legs (relaxed/ contracted), and state of arousal	Breathing patterns	0-1 points for each cue Crying earns 0-2 points Total score 0-7. ↑score indicates ↑ pain.
PAT	Hodgkin-son, Bear, Thorn, & Van Blaricum, 1994)	The PAT is a multidimensional scale created for clinical feasibility and *general pain*. It incorporates nurse perception of pain as 10% of the possible score. The focus of assessment is change over time.	posture/ tone, cry, sleep pattern, expression, color. Also: Nurse perception of pain	Respirations, heart rate, oxygen saturation, blood pressure	Scale is 0-20 points; all ten cues are worth 0-2 points.
PIPP	(Stevens, Johnston, Petryshen, & Taddio, 1996)	Developed as a multi-dimensional scale in full term and preterm infants for all types of pain. PIPP includes three facial cues from the NFCS that involve the top of the head	Facial actions: brow bulge, eye squeeze, nasolabial furrow Also, gestational age and behavioral state	heart rate, SaO_2	Scale is 0-21, 0-3 pts for each cue.

Table 4

Unidimensional Scale Psychometrics

Scale	Study and Sample	Validity			Reliability		
		Content	Construct*	Criterion	Interrater Reliability	Intrarater Reliability	Internal Consistency
ABC	(Bellieni et al., 2005) 90 FT neonates	yes	discriminant p<0.0001	Spearman *p*=0.91 with DAN	Cohen's k = 0.83	Cohen's k = 0.85 after 2 months	---
	(Bellieni et al., 2007a) 72 PT neonates	---	discriminant p<0.001	*r*=0.68 with PIPP	Cohen's k = 0.7	---	---
BIIP	(Holsti & Grunau, 2007b) 92 PT neonates	yes	p<0.0001	*r*=0.64 with NIPS	ICC=0.80-0.92	--	Cronbach's α=0.82
	(Holsti et al., 2008)	---	discriminant p<0.0001	---	ICC=0.79-0.92	---	---
CHIPPS	(Büttner & Finke, 2000) 584 newborns, infants and children	yes	Sensitivity= 0.92-0.96 Specificity = 0.74-0.95	Ø (used a toddler scale)	*r*=0.93	---	Cronbach's α=0.96 For infants
COMFORT-B	(van Dijk et al., 2000) 158 0-3 year olds, 80%<1year old	yes	---	0.89-0.96 with VAS	Cohen's K = 0.54-0.74	---	Cronbach's α= 0.90-0.92
	(Johansson & Kokinsky 2009) 40 infants and children, 87% <1 year old	yes	---	GCC= 0.50 with VAS	Cohen's kappa= 0.71	---	---
EDIN	(Debillon et al., 2001) 36-76 preterm infants	---	p<0.0001	---	Kappa= 0.59-0.74	---	Cronbach's α=0.86-0.94
FLACC	(Merkel et al., 1997) 89 children up to 7 yrs	yes	p<0.001	*r*=0.8 With OPS	*r*=0.94	---	---
	(Johansson & Kokinsky, 2009) 40 infants and children	yes	p<0.001	GCC= 0.50 with VAS	Cohen's kappa= 0.63	---	---
	(Ahn & Jun, 2007) 110 PT infants	---	discriminant p<0.001	*r*=0.82-0.84 with CRIES	Tested before assessment	---	---
NFCS	(Grunau & Craig, 1987) 77 FT neonates	yes	discriminant significant	---	0.88 (type unlisted)	---	---
	(Grunau et al., 1998) 40 PT infants, 5-56 days old	yes	discriminant p<0.0001	No value, correlate with HR	Cohen's kappa= 0.83	---	---
	(Pereira et al., 1999) 70 FT neonates	---	discriminant p<0.00001	---	---	---	---
	(Stevens, Johnston & Horton, 1994) 124 PT neonates	---	p<0.0001	---	---	Cohen's K= 0.90, after 2 weeks	---

Note.The first study listed for each scale is the original publication for the tool. Construct validity labeled *discriminant* indicates scales that cannot only distinguish between painful events and a baseline state, but can distinguish significantly between painful events and non-painful or stressful events. P and p-value refer to the likelihood that results could be duplicated by chance. *Spearman's p*, *r* (Pearson's r), *GCC* (gamma correlation coefficient), and *ICC* (intraclass correlation coefficient) are statistical values related to correlation that range in value from 1, indicating zero correlation, to 1, indicating absolute correlation. *Cohen's Kappa* is related to interrater agreement and also ranges from 0 to 1. *Cronbach's alpha* is a statistical measure of internal consistency ranging from 0 to 1.

Table 5

Multidimensional Scale Psychometrics

Scale	Study and Sample	Validity			Reliability		
		Content	Construct*	Criterion	Interrater	Intrarater	Internal Consistency
COM-FORT	(Ambuel et al., 1992) 37 PICU pts, 76% <2 years	Yes	---	*r*=0.75 with stress VAS	*r*=0.84	---	Cronbach's α=0.90
	(Wielenga et al., 2004) 19 PT infants	---	Sensitivity= 100%, specificity= 77%	*r*=0.84 with expert Likert	Weighted Kappa = 0.84	---	---
CRIES	(Krechel and Bildner, 1995) 24 post-op infants, 32-60 weeks GA	yes	discriminant (p<0.0001)	Spearman rank= 0.73 with OPS	*r*=0.72	---	---
	(Suraseranivongse et al., 2006) 22 FT post-op neonates	---	discriminant (p<0.001)	*r*=0.30-0.38 w/ CHIPPS, *r*=0.32-0.39 with NIPS	ICC= 0.98	---	---
NIPS	(Lawrence et al., 1993) 38 FT and PT infants	yes	(p<0.001)	*r*=0.53-0.84 with VAS	*r*=0.92-0.97	---	Cronbach's α= 0.87- 0.95
	(Williams, Khattak, Garza, & Lasky, 2009)	---	(p<0.001)	---	Cohen's kappa= 0.61-0.79	85-98%, retested instantly	---
	(Suraseranivongse et al., 2006) 22 FT post-op neonates	---	discriminant (p<0.001)	*r*=0.84-0.88 with CHIPPS and *r*=0.32-0.39 with CRIES	ICC=0.98	---	---
N-PASS	(Hummel, et al., 2008)	yes	(p<0.0001)	ICC, Spearman's rank=0.61-0.83 with PIPP	ICC=0.85 -0.95	---	Cronbach's α= 0.82-0.72 for pain, 0.89 for sedation
	(Hummel, Lawlor-Lean, &Weiss, 2009)	yes	discriminant (p<0.001)	Spearman rho correlation= 0.743 with PIPP	ICC= 0.79-0.93	Spearman's *p*= 0.87 after a week, 0.85 after a year	Cronbach's α = 0.84-0.89 for procedural pain
PAT	(Hodgkinson et al., 1994) 20 PT and FT neonates	yes	Reported with no p-value given	---	---	---	---
	(Spence et al., 2005)	---	---	*r*=0.76 with CRIES	ICC= 0.85	---	---
PIPP	(Stevens et al., 1996) 237 PT and FT neonates	yes	discriminant (p<0.02)	---	---	---	Standardized item α= 0.71
	Ballantyne et al., 1999 43 PT and FT neonates	yes	discriminant (p = 0.0001)		ICC= 0.93-0.96	ICC= 0.94-0.98	

Note.The first study listed for each scale is the original publication for the tool. Construct validity labeled *discriminant* indicates scales that cannot only distinguish between painful events and a baseline state, but can distinguish significantly between painful events and non-painful or stressful events. P and p-value refer to the likelihood that results could be duplicated by chance. *Spearman's p*, *r* (Pearson's r), *GCC* (gamma correlation coefficient), and *ICC* (intraclass correlation coefficient) are statistical values related to correlation that range in value from

1, indicating zero correlation, to 1, indicating absolute correlation. *Cohen's Kappa* is related to interrater agreement and also ranges from 0 to 1. *Cronbach's alpha* is a statistical measure of internal consistency ranging from 0 to 1.

can be divided into procedural pain or prolonged pain. Procedural pain exists after a precise nociceptive event and is the easiest to quantify because the painful event can easily be controlled. Prolonged pain can be the result of a procedure, but manifests in a longer and more variable pain and recovery period, which is harder to assess. Researchers have developed pain tools for procedural pain, prolonged pain, preterm infants, full term infants, and infants on respirators (Anand, 2007). Additionally, several pain tools are intended for assessing stress in the infant as well, and researchers have developed several tools solely for evaluating stress. Tables 2 and 3 summarize some of the most widely used infant pain and stress assessment tools.

In conclusion, pain and stress assessment and resolution in infants is vital in the clinical setting. Unmanaged infant pain and stress is a problem with severe consequences, which can often be solved with pharmacological and/or non-pharmacological intervention. However, administration of analgesia within the infant population requires validation and awareness that the infant is actually in pain. With so many pain and stress assessment tools available for clinicians and researchers, selecting the most appropriate and most accurate pain assessment tool for the patient is paramount. Additionally, some assessment tools include so many scored items that they are only practical in a research setting, and are too complex for a clinical setting. The available pain assessment tools have been compiled and summarized previously (Duhn & Medves, 2004), but new research about pain assessment skills has been produced since that publication. Also, Duhn and Medves did not explore the assessment of stress in their systematic review. The purpose of this review is to (1) compare infant pain and stress instruments and make recommendations about the reliability, validity, and clinical feasibility of current pain and stress tools in both the research and clinical setting, (2) make recommendations about validation study

design to best prove reliability and validity, and (3) discuss the implications for researchers and clinicians regarding the most appropriate cues for assessing pain and stress in preterm infants to allow for appropriate pharmacological and non-pharmacological intervention.

Chapter 3: Methods

A complete and inclusive analysis of current pain and stress assessment tools requires an organized exploration of the literature. In this chapter, the methods used to select articles and tools for analysis will be described.

A critical review of the literature was conducted using the search terms *pain scales* or *measures of stress* in the PubMed database of the National Institutes of Health. 687 articles were found and 16 were selected for inclusion. 671 articles were rejected based on title and/or abstract. Within the PubMed search, limits were set to include articles about human infants (birth to 23 months) published in English since October 1999. The types of literature searched were limited to clinical trials, meta-analysis, randomized controlled trials and review articles. All studies were published in peer-reviewed, scientific journals. Studies were accepted based on the inclusion and validation of an infant pain or stress assessment tool and an average sample age of one to thirty days. Samples with both healthy and ill neonates were accepted.

After further examination of the selected articles, studies published before 2005 were excluded to capture more current research. Six of the sixteen previously selected articles were excluded due to the sample population age, poor study design, and publication before 2005. The bibliographies of the initial sixteen selected articles were searched for additional applicable tools and studies. All seventeen articles found within the bibliography search were rejected due to sample age or publication date, but were read to find additional pain tools. Two additional tools were identified, as listed in Table 7.

Ten studies and sixteen assessment tools were identified through PubMed and a bibliography review. The text *Pain in Neonates* (Anand, Stevens, & McGrath, 2007) was consulted to find additional pain tools and thirteen more tools were found for study. An

individual PubMed search was performed for each of the twenty-nine tools with limitations and exclusion criteria identical to the previous search. Nine additional articles were found for review. The selection process is described in Figure 1.

Out of the twenty-nine pain or stress assessment tools searched, thirteen were selected for critical review and analysis. Tables 6, 7, and 8 list the twenty-nine original tools. This decision to include certain tools was based on the presence of recent validation research. The psychometric data for the selected tools are listed in Tables 4 and 5.

Figure 1. Description of Article Selection Process

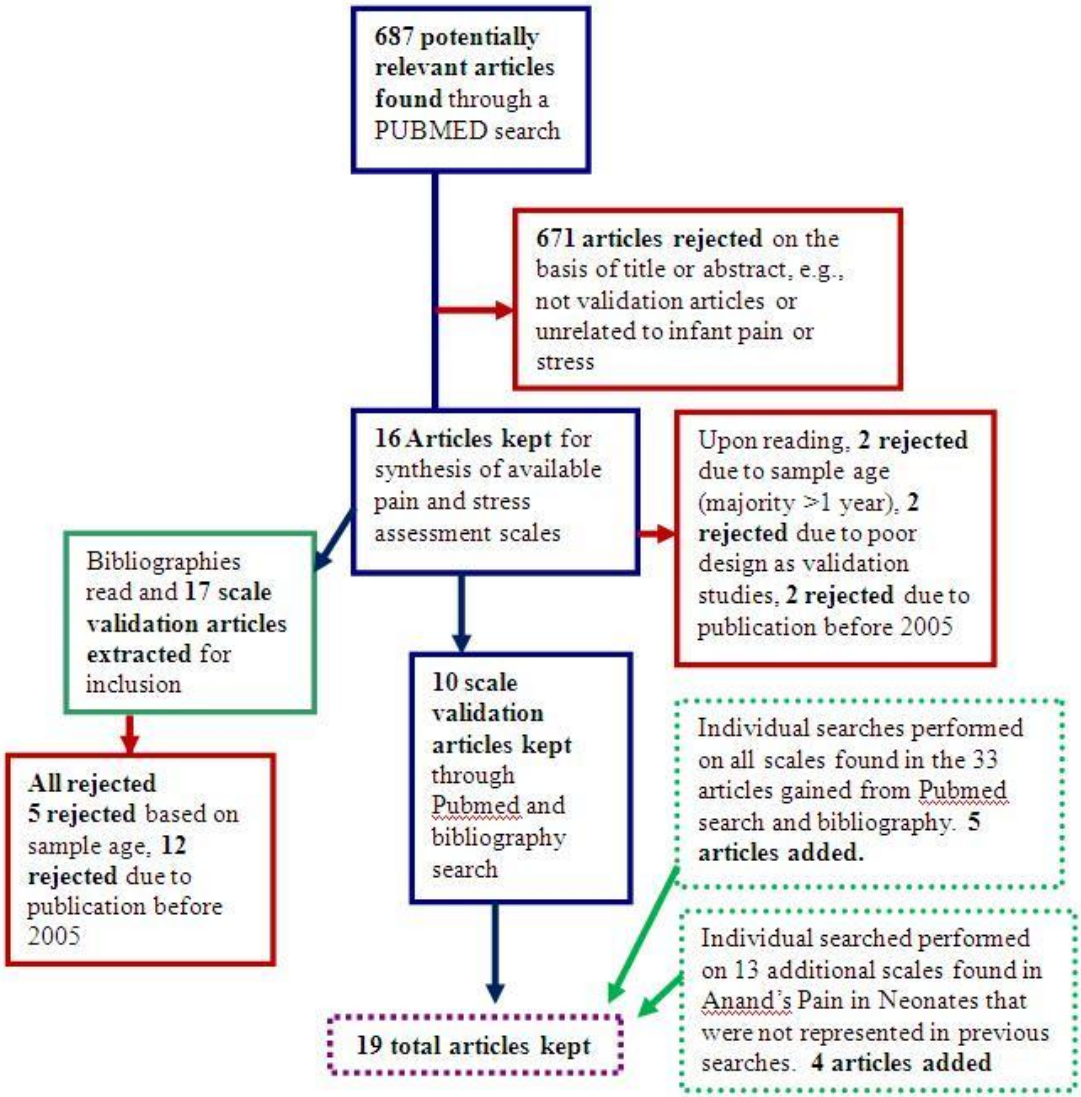

Table 6

Scales Represented in PubMed Using Search Terms Measures of Stress *and* Pain Scales

Scale Name	Full Name
ABC Scale*	Acuteness of the first cry, Burst rhythmicity, Constancy of cry
BIIP Scale*	Behavioral Indicators of Infant Pain
CHIPPS*	The Children's and Infants' Postoperative Pain Scale
COMFORT scale*	COMFORT scale
CRIES scale*	Crying, Requires increased oxygen administration, Increased vital signs, Expression, Sleeplessness.
DAN scale	Douleur Aiguë Nouveau-né
FLACC scale*	Face, Legs, Activity, Cry, Consolability
MBPS	Modified Behavioral Pain Scale
NFCS*	Neonatal Facial Coding System
NIPS*	Neonatal Infant Pain Scale
NNNS	NICU Network Neurobehavioral Scale
PIPP*	Premature Infant Pain Profile
PAIN	Pain Assessment in Neonates

*Scales were selected for inclusion and further evaluation within this review

Table 7

Scales Represented in a Bibliography Search of Preselected Articles

Scale Name	Full Name
DSVNI scale	Distress Scale for Ventilated Newborn Infants
SUN scale	Scale for Use in Newborns

Table 8

Scales Represented in Anand, Stevens, & McGrath (2007)

Scale Name	Full Name
DSVNI scale	Distress Scale for Ventilated Newborn Infants
SUN scale	Scale for Use in Newborns
BFACS	Baby Facial Action Coding System
BPS	Behavioral Pain Score
BPS	Bernese Pain Scale
CAAS	Cardiac Analgesic Assessment Scale
EDIN Scale*	Échelle Douleur Inconfort Nouveau-né
IBCS	Infant Body Coding System
LIDS	Liverpool Infant Distress Scale
MAX	Maximally Discriminative Facial Movement Coding
MAPS	Multidimensional Assessment of Pain Scale
N-PASS*	Neonatal Pain, Agitation, and Sedation Scale
NNICUPAT	Nepean NICU Pain Assessment Tool
PAT*	Pain Assessment Tool
RIPS	Riley Infant Pain Scale

*Scales were selected for inclusion and further evaluation within this review

The nineteen studies and thirteen scales picked for inclusion represent the most current studies validating the most researched infant pain and stress assessment tools available. The aims of this integrated review of the literature are to (1) analyze the tools based on the psychometrics available in these studies, (2) to make recommendations about their utility in research and clinical settings, and (3) to summarize the research concerning the assessment of stress and pain in the neonate.

Chapter 4: Results

The data compiled from the analysis of the selected literature are reviewed in this section. The sample statistics, study design, scale psychometrics, and study qualities contributing to validity and reliability are described within these results. Table 9 summarizes the studies analyzed in this integrative review.

Table 9

Summary of Studies Reviewed

#	Author group	Scales assessed	Age group average (range)	Sample characteristics	Type of pain	Study design
1	(Bellieni et al., 2005)	ABC	Assessed at 0-3 days GA: 39 wks (38-41)	90 infants in an Italian hospital, no intubation	Procedural - heel prep, heel prick,	Part RCT, part cohort
2	(Bellieni et al., 2007a)	ABC	GA: 34.1±2.7 wks	72 PT infants in an Italian hospital, no intubation	Procedural – heel prick and cotton swab to heel	Part RCT, part cohort
3	(Holsti & Grunau, 2007b)	BIIP	Assessed at 25-28 wks and/or 32 weeks PCA GA@B: 29 wks (24-32)	92 PT infants, some intubated, from a NICU in British Columbia.	Procedural - heel prick	repeated measures cohort study
4	(Holsti et al., 2008)	BIIP	PNA at assessment: 21 days (3-59) GA@B: 24-32 wks	69 PT infants, 20% intubated, from a NICU in British Columbia	Procedural and routine care (diaper change)	Within subjects, repeated measures cross over study
5	(Suraserani vongse et al., 2006)	CHIPPS CRIES NIPS	PNA at assessment: 1 day (1-23 days) GA: 39.9 wks	22 ft neonates, 50% intubated in a Thai hospital.	Post op - prolonged	prospective observational study, part cohort
6	(Johan-sson & Kokinsky, 2009)	COM-FORT-B FLACC	4 months (0-108m) 87% are <1 yr old, prematurity at birth not mentioned	40 post-op intubated PICU patients in a Swedish hospital	Post op - prolonged	Prospective observational, part cohort
7	(Wie-lenga et al., 2004)	COMFORT	PNA at assessment: 0-4 days GA: 30 wks ($26\text{-}36^{4}$)	n=19 PT ventilated pt infants from a NICU in Amsterdam	Post op - sedation	Sample of an open population, prospective observational
8	(Ahn & Jun, 2007)	CRIES FLACC PIPP	<7 days old Average GA@B: 32^{3} wks ($25^{6}\text{-}41^{6}$)	110 PT infants from a level III NICU in Korea, intubated pts included	Procedural (3 levels of pain/stress intensity)	Exploratory correlational
9	(Ancora et al 2009)	EDIN	GA: 30 wks ($25^{1}\text{-}41^{2}$). Exact PNA at assessment not reported	84 PT newborns in an Italian NICU, 21% mech ventilation	Prolonged pain	Retrospective observational

10	(Serpa, 2007)	NFCS NIPS PIPP	1-7 days, GA 27-36 wks	n=11 pt infants assessed at 1, 5, and 7 days PNA in a Brazilian NICU, intubated patients included	Procedural - venepuncture	Prospective cohort study
11	(Humme et al., 2009)	N-PASS	1-30 days, GA@B 23-40 wks	N=42 pt and ft infants in two Illinois NICUs, 63% ventilated/CPAP.	Procedural – real vs sham heel stick	Prospective crossover design.
12	(Taksande et al 2005)	NIPS	3.6/3.7 days (0-7) GA 37-42 wks	n=80 neonates split into one group >2.5 kg and another <2.5 kg from a hospital in India. Intubation not mentioned.	Procedural - venepuncture	prospective cohort study
13	(Bellieni, Cordelli, et al 2007)	PIPP NIPS	2 groups: (1) GA:34.2 (2) GA: 39.5	N=20 PT and 20 FT infants in an Italian hospital. No ventilation info given	Procedural - heel stick	prospective cohort study
14	(Hummel, et al., 2008)	N-PASS	0-100 days GA@B: 23-40 wks	n=46 intubated and/or postsurgical infants on a Level III NICU in Illinois.	Prolonged – analgesia measured	Prospective observational
15	(Spence et al., 2005)	PAT	22 days (0-182) GA@B: 36.1 wks	n=144 infants from two Australian NICUs. 45% were pt, 55% ft. 33% ventilated.	General pain – evaluated for whole shift.	Prospective observational
16	(Jons-dottir & Kristjans-dottier, 2005)	PIPP	9 days (0.7-28 GA@B: 33.9 wks (26.3-39.5)	n= 24 PT and FT infants from a level III NICU in Iceland. 17% on CPAP	Procedural Pain – heel lance and diaper change	Crossover design (non-randomized)
17	(Eriksson et al., 2008)	PIPP	3-5 days GA: 40.2 wks	n= 27 healthy ft infants from a maternity unit in Sweden	Procedural – heel lance and tactile stimulation	Crossover design
18	(Vederhus et al., 2006)	PIPP	0-30 days, GA at birth: 59% <36 wks, 41% >36 wks	n=111 PT and FT consecutive neonates from a NICU and maternity ward in Norway, intubated and ventilated patients included.	Procedural – diaper change and heel stick/ intravenous cannula insertion	Prospective cohort - known groups comparison design
19	(Slater et al., 2008)	PIPP	24-34.6 GA@B, 5-134 days old	n=12 PT infants from a neonatal unit in London, intubated and ventilated patients included.	Procedural – heel lance	Prospective cohort (not observational)

Study Sample

Location.

Nineteen studies were selected for inclusion within the review. The studies originated from five continents: Europe, Asia, North America, Australia, and South America. The majority of settings were neonatal intensive care or surgical intensive care units; however, several studies

did not indicate the specific hospital setting in which they were conducted (Bellieni et al., 2005; Bellieni et al., 2007a; Bellieni et al., 2007b). One study was conducted in a pediatric intensive care unit (Johansson & Kokinsky, 2009), and another in a maternity unit (Eriksson et al., 2008).

Sample characteristics.

Each of the nineteen study samples was comprised of infants less than one year of age, except for one study in which 87% of the sample was comprised of infants less than one year of age, while the remaining 13% of the sample included children from one to ten years of age (Johansson & Kokinsky, 2009). Reporting gestational age and postnatal age was inconsistent among researchers; gestational age is a term than can represent the age of an infant at birth, and is often but not always reported as *gestational age at birth*. Researchers also occasionally use the term *gestational age* (GA) to represent a neonate at their current post-conceptual age at assessment. This is a less accurate representation of true age, as it does not account for development that occurs post birth. An adjusted gestational age uses the gestational age and postnatal age since birth. This is the best representation of an infant's true developmental age. In the nineteen selected studies, several researchers reported the gestational age of the infants without reporting the postnatal age and without clarifying whether the gestational age represented the age of the infant at birth or at the time of assessment (Bellieni et al., 2007a; Bellieni et al., 2007b). Some researchers reported both a gestational age and a postnatal age, but it was still not clear whether the gestational age reported represented the infants' age at birth or at assessment. For these cases, the average subject's postnatal age was less than five days (Eriksson et al., 2008; Suraseranivongse et al., 2006; Taksande et al., 2005), so the difference may not be clinically important.

Gestational age.

The studies varied in the gestational age (GA) at birth of the samples. Five studies used a strictly full term sample, defined as infants born at 37 to 42 weeks gestation (Bellieni et al., 2005; Eriksson et al., 2008; Johansson & Kokinsky, 2009; Suraseranivongse et al., 2006; Taksande et al., 2005), while seven employed a completely preterm sample, defined as less than 37 weeks gestation (Bellieni et al., 2007a; Holsti et al., 2008; Holsti & Grunau, 2007b; Serpa et al., 2007; Slater et al., 2008; Wielenga et al., 2004). Seven studies used a mixed sample of term and preterm infants (Ahn & Jun, 2007; Ancora et al., 2009; Bellieni et al., 2007b; Hummel et al., 2008; Hummel et al., 2009; Jonsdottir & Kristjansdottir, 2005; Spence et al., 2005; Vederhus et al., 2006). In these studies, comparisons were made between scale composite scores for preterm and full term infant groups.

Effect of gestational age on pain scores.

Five of the studies with mixed gestational age samples examined the effect of gestational age on pain scores. In a 2009 publication validating the N-PASS, Hummel and colleagues found that the pain scores of infants who were <30 weeks GA were significantly lower than the scores of infants older than 30 weeks GA experiencing the same painful procedure ($p<0.012$) (Hummel et al., 2009). Researchers validating the EDIN tool for prolonged pain (0-15 point scale) found through a retrospective chart review that full term infants earned average scores of 5, infants 32-37 weeks GA earned average scores of 4 and infants 25-31 weeks GA earned average scores of 3 in a NICU setting ($p<0.001$) (Ancora et al., 2009). In a study by Ahn and Jun comparing the CRIES, FLACC and PIPP tools during a painful event, preterm infants were rated lower than full term infants using the CRIES tool (scores of 4.62 v 5.73, $p<0.05$) and FLACC tool (scores of 4.26 v 6.09, $p<0.05$). No significant difference in pain scores was seen between preterm and full

term infants assessed with PIPP, a tool which assigns extra points to infants born prematurely (Ahn and Jun, 2007). Alternatively, another study validating the PIPP demonstrated much higher scores in severely preterm infants after adjusting for gestational age, but the researchers reported the sample size was too small to state significance (Vederhus et al., 2006). A study validating the PAT tool did not show a difference between preterm and full term infant pain scores (Spence et al., 2005). In addition to variations in gestational age, the studies also differed according to the postnatal age of the infants at assessment.

Postnatal age.

Within the studies that reported postnatal age, samples ranged from strictly neonates to average sample age of four months. Ten studies used a strictly neonatal sample, defined as 0-28 days old (Ahn & Jun, 2007: Bellieni et al., 2005; Eriksson et al., 2008; Hummel, Lawlor-Lean, & Weiss, 2009; Jonsdottir & Kristjansdottir, 2005; Serpa et al., 2007; Suraseranivongse et al., 2006; Taksande et al., 2005; Vederhus et al., 2006; Wielenga et al., 2004.). Three more studies featured samples in which the average age of subjects was less than 30 days, but the range included infants up to six months of age (Holsti & Grunau, 2007b; Holsti et al., 2008; Slater et al., 2008; Spence et al., 2005). One study featured a sample with an average age of 35 days, and two studies failed to report an average sample age (Ancora et al., 2009; Hummel et al., 2008). Two studies did not report postnatal age, but referred to subjects as neonates (Bellieni et al., 2007a; Bellieni et al., 2007b), and one study used a sample with an average age of four months (Johansson & Kokinsky, 2009).

Inclusion/exclusion criteria.

In order to be assessed with a pain tool, patients are required to be able to mount a behavioral or physiological response to pain. Thus, in selecting sample criteria, researchers often

exclude patients whose pain cannot be appropriately measured due to confounding factors such as intubation, sedation, and congenital and neurological anomalies.

Two of the 19 studies excluded patients who were mechanically ventilated or intubated. These studies validated the *Acuteness of the first cry, Burst rhythmicity and temporal Constancy of cry intensity* scale (ABC scale), a tool which assesses cry characteristics and cannot be performed accurately with an intubated subject (Bellieni et al., 2005; Bellieni et al., 2007a). Three studies did not discuss intubation and mechanical ventilation in their exclusion criteria or their results (Bellieni et al., 2007b; Eriksson et al., 2008; Taksande et al., 2005). Of the fourteen studies that included intubated patients, some used validated tools that included an item assessing *cry*. The NIPS, PAT, CRIES, CHIPPS, COMFORT-B, and FLACC scales all include a cry cue and all were applied to ventilated infants in at least one study within this review. To use and validate the tools accurately, Johansson and Kokinsky used modified versions of the Face, Legs, Activity, Cry, and Consolability scale (FLACC scale) and the COMFORT-B scale in their 2009 study of intubated infants and children. The researchers assessed for *cry face* instead of *cry* in relation to ventilated infants (Johansson & Kokinsky, 2009). Johansson and Kokinsky describe infants with cry face as silently moaning or whimpering, with facial expressions that mimic screaming or sobbing. The Neonatal Facial Coding System also involves facial features that typify cry face, which include neonatal brow bulge, eye squeeze, nasolabial furrow, and open lips (Serpa et al., 2007).

Seventeen of the nineteen studies excluded infants with neurological or congenital anomalies that could possibly alter pain response. Two studies accepted infants regardless of their neurological status (Suraseranivongse et al., 2006; Spence et al., 2005). Seven studies excluded infants who had received sedatives or analgesia within 72 hours of assessment, and

three excluded infants who had received any neuromuscular blockage or muscle relaxers. Of the nine studies that put no limits on infants who had received analgesia or sedation, three had healthy, full term samples that would not likely require extensive sedation or analgesia (Bellieni et al., 2005; Eriksson et al., 2008; Taksande et al., 2005), two had convenience samples in which every consecutively born infant admitted to a NICU was assessed without exception (Ancora et al., 2009; Bellieni et al., 2007b), and two validated the COMFORT scale and N-PASS, which both assess sedation in addition to pain (Hummel et al., 2008; Wielenga et al., 2004).

Study Design

All nineteen studies were quantitative and eighteen of the nineteen were prospective, meaning the subjects were recruited before the events that were studied occurred, thus participants were studied before and after the procedural events. One study featured a retrospective chart review (Ancora et al., 2009). All studies were either observational, in which the infants were assessed without intervention, or quasi-experimental, in which infants' reactions to an event were studied. There were basic differences between the aims, methods, and types of pain assessed in the observational and quasi-experimental studies.

Observational studies.

Most of the observational studies were associated with postoperative pain or sedation and involved the Neonatal Pain, Agitation and Sedation Score (N-PASS), COMFORT scale, COMFORT behavioral scale (COMFORT-B scale), Echelle Douleur Inconfort Nouveau-Ne (EDIN scale), Crying, Requires oxygen, Increased vital signs, Expression and Sleeplessness (CRIES scale), Pain Assessment Tool (PAT) and Children and Infants Postoperative Pain Scale (CHIPPS) (Ancora et al., 2009; Hummel et al., 2008; Johansson & Kokinsky, 2009; Spence et al., 2005; Suraseranivongse et al., 2006; Wielenga et al., 2004). All of these scales were created

to assess non-procedural pain, like postoperative pain, acute-prolonged pain, general pain, and sedation. One procedural scale, NIPS, was also tested for post-operative pain (Suraseranivongse et al., 2006). Within these studies, pain assessment did not occur with any one event, but instead pain or sedation was often assessed periodically in accordance to individual hospital unit policy. Some of these studies involved an element of experimental design – two assessed pain before and after administering analgesia to measure the construct validity of FLACC and N-PASS (Johansson & Kokinsky, 2009; Hummel et al., 2008), or used an additional sample of subjects to assess pain during a procedural event (Suraseranivongse et al., 2006).

Quasi-experimental studies.

Most of the quasi-experimental studies within this review assessed procedural pain. These studies featured a specific painful event, and the subjects were assessed with the given scales before, during, and/or after the event (Bellieni et al., 2005; Bellieni et al., 2007a; Bellieni et al., 2007b, Holsti & Grunau, 2007b; Serpa et al., 2007; Slater et al., 2008; Taksande et al., 2005). A painful event was usually invasive and defined as a heel stick or heel lance, venepuncture, or intravenous catheter insertion, all of which were medically necessary procedures. Other quasi-experimental studies explored the effects of two separate events, a painful event and a stressful or non-pain event to determine the ability of a tool to discriminate between pain and non-pain stimulation. The PIPP, BIIP, and N-PASS tools were assessed with this type of intervention. The time between pain and non-pain/stress events varied for each study; for some, the procedures were conducted in a randomized order, with five to ten minutes between events (Eriksson et al., 2008; Hummel, Lawlor-Lean, & Weiss, 2009), or in a specified order with approximately six minutes between events (Jonsdottir & Kristjansdottir, 2005). Other researchers did not perform the events during a specific timeframe; assessment was simply

conducted when the infant required a pain or non-pain procedure (Holsti et al., 2008; Vederhus et al., 2006). This approach was used for a study validating CRIES, FLACC, and PIPP which assessed eight different events, both pain and stress-inducing, divided into three levels of stimulation: invasive procedures, cluster care, and auditory stimuli (Ahn & Jun, 2007).

Tool Psychometrics

The tools tested within the studies in this review were validated for their psychometric properties. The studies were validated for pain assessment, not just by the procedures they assessed, but by the demographics of the raters performing the assessments, the context of the rating, and the states of pain and stress that the tools could discern. Psychometric data for the studies reviewed are presented in Tables 10 and 11.

Construct validity.

Most of the quasi-experimental studies included tests for construct validity through assessment of pain with the scale in question before, during, and after a painful event. In the case of some observational studies, construct validity was found by assessing for pain before and after analgesia administration.

Discriminant validity, sensitivity, and specificity.

In addition to construct validity, some of the studies under review reported discriminant validity, sensitivity and specificity. Discriminant validity is the ability of a tool to distinguish between two separate theoretical concepts, such as *pain* and *non-pain* or *stress*. Sensitivity and specificity are two additional forms of construct validity related to discriminant validity. Sensitivity is the ability of a tool to detect the presence of a construct, or determine when an

Table 10

Psychometrics for Unidimensional Scales Reviewed

Scale	Study and Sample	Validity			Reliability		
		Content	Construct*	Criterion	Interrater Reliability	Intrarater Reliability	Internal Consistency
ABC	(Bellieni et al., 2005) 90 FT (full term) Neonates **Nurse rated feasibility**: *good*	yes	discriminant p<0.0001	Spear-man *p*=0.91 with DAN	Cohen's k = 0.83	Cohen's k = 0.85 after 2 months	---
	(Bellieni et al., 2007a) 72 PT (preterm) neonates **Nurse rated feasibility**: *good* and *very good*	---	discriminant p<0.001	*r*=0.68 with PIPP	Cohen's k = 0.7	---	---
BIIP	(Holsti & Grunau, 2007b) 92 PT neonates	yes	p<0.0001	*r*=0.64 with NIPS	ICC=0.80-0.92	---	Cronbach's α=0.82
	(Holsti et al., 2008) 69 PT neonates	---	discriminant P<0.0001	---	ICC=0.79-0.92	---	---
CHIPPS	(Suraseranivongse et al., 2006) 22 FT post-op neonates **Nurse rated feasibility**: Practicality rated 4.5-8.5/10	---	Discriminant p<0.001	*r*=0.30-0.38 with CRIES, *r*=0.84-0.88 with NIPS	ICC=0.93	---	---
COM-FORT-B	(Johansson & Kokinsky, 2009) 40 infants and children 87% <1 year old	Yes	---	GCC= 0.50 with VAS for pain, 0.57 with NIS for sedation	Cohen's kappa: 0.71	---	---
EDIN	(Ancora et al., 2009) 84 PT infants	---	---	---	---	---	---
FLACC	(Johansson & Kokinsky, 2009) 40 infants and children	yes	p<0.001	GCC= 0.50 with VAS	Cohen's kappa: 0.63	---	---
	(Ahn & Jun, 2007) 110 PT infants	---	discriminant p<0.001	*r*=0.82-0.84 with CRIES	Coders checked before rating	---	---
NFCS	Serpa et al., 2009 11 PT neonates	---	Yes, no p-value	---	---	---	---

Note:Construct validity labeled *discriminant* indicates scales that cannot only distinguish between painful events and baseline, but can distinguish significantly between painful events and non-painful or stressful events. P and p-value refer to the likelihood that results could be duplicated by chance. *Spearman's p*, *r* (Pearson's r), *GCC* (gamma correlation coefficient), and *ICC* (intraclass correlation coefficient) are statistical values related to correlation that range in value from 1, indicating zero correlation, to 1, indicating absolute correlation. *Cohen's Kappa* is related to interrater agreement and also ranges from 0 to 1. *Cronbach's alpha* is a statistical measure of internal consistency ranging from 0 to 1.

Table 11

Psychometrics for Multidimensional Scales Reviewed

Scale	Study and Sample	Validity			Reliability		
		Content	Construct	Criterion	Interrater	Intrarater	Internal Consistency
COM-FORT scale	Wielenga et al., 2004 19 PT infants	---	Sensitivity: 100%, specificity: 77%	*r*=0.84 with expert Likert	Weighted Kappa = 0.84	---	---
CRIES	(Suraseranivongse et al., 2006) 22 FT post-op neonates	---	discriminant p<0.001	*r*=0.30-0.38 with CHIPPS, *r*=0.32-0.39 with NIPS	ICC: 0.98	---	---
	Ahn & Jun 110 PT infants **Nurse rated feasibility**: Practicality rated 7.5-8.0/10	---	---	*r*=0.899 with FLACC and 0.601 with PIPP	Coders checked before assessment	---	---
NIPS	Serpa et al., 2009 11 PT neonates	---	Yes, p-value not given	---	---	---	---
	(Taksande et al., 2005) 80 FT neonates	---	p<0.001	---	---	---	---
	(Suraseranivongse et al., 2006) 22 FT post-op neonates **Nurse rated feasibility**: Practicality rated 8.7-9.1/10	---	discriminant p<0.001	*r*=0.84-0.88 with CHIPPS and *r*=0.32-0.39 with CRIES	ICC=0.98	---	---
	(Bellieni et al., 2007b) 40 PT and FT neonates	---	---	---	Cohen's K of 0.3-0.6	---	---
N-PASS	(Hummel, et al., 2008) 46 PT and FT post-op neonates	yes	(p<0.0001)	ICC, Spearman's rank=0.61-0.83 with PIPP	ICC=0.85-0.95	---	Cronbach's α= 0.82-0.72 for pain, 0.89 for sedation
	(Hummel, Lawlor-Lean &Weiss, 2009) 44 PT and FT post-op neonates	yes	discriminant P<0.001	Spearman rho correlation: 0.743 with PIPP	ICC= 0.79-0.93	Spear-man's *p*= 0.87 after 1 week, 0.85 after one year	Cronbach's α = 0.84-0.89 for procedural pain
PAT	(Spence et al., 2005) 144 FT and PT neonates and infants	---	---	Pearson's *r*=0.76 with CRIES	ICC= 0.85	---	---
PIPP	(Ahn & Jun, 2007) 110 PT infants	---	---	*r*=0.653 with FLACC and *r*=0.601 with CRIES	Coders checked before assessment	---	---
	(Jonsdottir et al., 2005) 24 PT and FT neonates	yes	discriminant p<0.0001	--	*r*=0.887 - 0.961	---	---
	(Eriksson et al., 2008) 27 FT neonates	---	Discriminant p=0.004	---	---	---	---

(Vederhus et al., 2006) 111 PT and FT neonates	yes	discriminant p=0.014	---	Spearman correlation coefficient 0.89 -0.97	---	Cronbach's alpha = 0.78
(Slater et al., 2008) 12 PT infants and neonates	---	---	regression coefficient= 0.72 with cortical activity level	ICC n = 0.96	Intraclass correlation = 0.99	---
(Bellieni et al., 2007b) 40 PT and FT neonates	---	---	---	Cohen's k=0.1-0.16	---	---

*Construct validity labeled *discriminant* indicates scales that cannot only distinguish between painful events and baseline, but can distinguish significantly between painful events and non-painful or stressful events. P and p-value refer to the likelihood that results could be duplicated by chance. *Spearman's p*, *r* (Pearson's r), *GCC* (gamma correlation coefficient), and *ICC* (intraclass correlation coefficient) are statistical values related to correlation that range in value from 1, indicating zero correlation, to 1, indicating absolute correlation. *Cohen's Kappa* is related to interrater agreement and also ranges from 0 to 1. *Cronbach's alpha* is a statistical measure of internal consistency ranging from 0 to 1.

infant is in pain, while specificity is the ability of a tool to detect the absence of a construct, or determine when an infant is not in pain.

Within the reviewed studies, nine tested discriminant validity or sensitivity and specificity of pain tools between the constructs of *pain* and *non-pain* or *pain* and *stress* by comparing the scores of infants at both a painful event and either a *non-pain* or *stressful* event.

Five of these studies compared pain to a non-pain event; non-pain events included a cotton swab to the heel (Bellieni et al., 2007a), thirty seconds of light tactile stimulation to the infants' arm (Eriksson et al., 2008), and the preparatory activities for a painful event (Bellieni et al., 2005; Hummel et al., 2009; Suraseranivongse et al., 2006). Four of these studies compared painful events to a more stressful event; two assessed diaper change (Jonsdottir & Kristjansdottir, 2005; Vederhus et al., 2006), and two assessed various types of cluster care, like diaper change, temperature, mouth care, etc (Ahn & Jun, 2007; Holsti et al., 2008).

Four of the nine studies reported a significant difference between a *pain* event and a *non-pain* or *stress* event without labeling as any type of construct validity (Ahn & Jun, 2009; Eriksson et al., 2008; Jonsdottir & Kristjansdottir, 2005; Vederhus et al., 2006), two reported it as

discriminant validity (Holsti et al., 2008; Hummel et al., 2009) and three reported it as sensitivity and specificity (Bellieni, 2005; Bellieni et al., 2007a; Suraseranivongse et al., 2006).

Criterion validity.

Ten of the nineteen studies reviewed assessed a pain tool for criterion validity, which is the level of agreement of a scale with a *gold-standard* that has previously been validated for assessing pain. The tools validated in the reviewed studies were often compared to tools that, while measuring the same construct of pain or stress, were fundamentally different. Three studies compared multidimensional tools with unidimensional tools (Bellieni et al., 2007a) (Holsti & Grunau, 2007b; Suraseranivongse et al., 2006) and two studies compared tools with behavioral or physiological cues to Likert-style scales (Johansson & Kokinsky, 2009; Wielenga et al., 2004). In five studies, tools being validated were compared to similar tools; behavioral tools were compared to behavioral, multidimensional compared to multidimensional (Bellieni et al., 2005; Hummel et al., 2008; Hummel et al., 2009; Johansson & Kokinsky, 2009; Spence et al., 2005).

In one study, NIPS is referred to as a strictly behavioral scale (Suraseranivongse et al., 2006) while in another it is considered multidimensional (Holsti & Grunau, 2007b). The item which Holsti and colleagues consider physiological is *breathing pattern,* which can be scored as *relaxed* (0 points) or *change in breathing* (1 point). It is the opinion of the current author that the NIPS scale is multidimensional.

Pain and stress definition.

Certain studies defined cut-off scores for pain for each scale. Jonsdottir and Kristjansdottir (2005) defined moderate to severe pain as a score greater than 11.72 points out of a possible 21 points, and minimal to no pain as a score less than 6 points. They noted that these

numbers corresponded to previous studies for the PIPP which defined scores of 12 and 6 as the pain and non-pain limits. In another study, pain requiring intervention was defined as a score greater than 4 using the CRIES, CHIPPS, and NIPS scales. CRIES and CHIPPS have a 0-10 scoring scale, while NIPS has a 0-7 scale (Suraseranivongse et al., 2006).

Sedation.

Two of the tools within this review assess for both pain and sedation. The COMFORT scale and the N-PASS are two tools which are designed to assess whether a patient is very sedated, has intense pain, or is somewhere in between the two. A score of 0 on the N-PASS indicated an absence of both pain and sedation.

Study Qualities Contributing to Validity and Reliability

Rater demographics.

In validating a tool, a practical way to select those who assess subjects within a study is to evaluate who would use the tool in the setting for which it is designed. In many healthcare settings, the caregiver most often responsible for pain assessment is a nurse. As a result, many studies researching clinical feasibility assign nurses to administer the tools. For studies examining the tool for use in a research setting, the nurse rater is not as necessary. Many researchers also used more than one rater, which allows the tool to be tested for interrater reliability. Interrater reliability is determined by comparing the scores of two separate raters. If a tool is reliable between raters, different people using the tool to assess the same situation should produce a similar score.

The studies within this review used a variety of raters. Four of the studies within this review used two nurses or two physicians to assess pain with the ABC scale (Bellieni et al.,

2005; Bellieni et al., 2007a), COMFORT-B scale, FLACC scale (Johansson & Kokinsky, 2009), and the N-PASS (Hummel et al., 2008). Two studies used three nurses at each assessment; one validated CHIPPS, CRIES, and NIPS (Suraseranivongse et al., 2006), and the other NIPS and PIPP (Bellieni et al., 2007b). Some researchers paired an expert in the subject with a novice, to reflect potential raters with varied experience with scoring pain. Five studies used either a nurse and an expert or an expert and a non-expert, for BIIP, COMFORT, N-PASS, and PIPP (Holsti & Grunau, 2007b; Holsti, Grunau, Oberlander, & Osiovich, 2008; Hummel et al., 2009; Jonsdottir & Kristjansdottir, 2005; Wielenga et al., 2004). Five studies used one rater at each assessment, which involved the EDIN scale (Ancora et al., 2009), NFCS, NIPS, and PIPP (Eriksson et al., 2008; Serpa et al., 2007; Taksande et al., 2005; Vederhus et al., 2006). Two studies used raters who were identified as researchers and research assistants for the CRIES, FLACC, and PIPP (Ahn & Jun, 2007; Slater et al., 2008).

Training process for raters.

Few researchers described the training process for the use of each tool within their studies, but those that did usually reported that the process was brief. Training with the cry-based ABC scale involved two 20 minute sessions with an audio compact disc (Bellieni et al., 2005). Wielenga et al., 2004 describe a two hour training process for the COMFORT scale for the nurses within the study. Ancora and colleagues did not define the training time for the EDIN scale, but reported that all nurses and neonatologists working on the unit where the study took place had to attend a course in neonatal pain (Ancora et al., 2008). An introduction to the PIPP scale for a NICU nurse without experience in pain assessment was described as *brief* (Jonsdottir & Kristjansdottir, 2005). All other studies either did not mention training or explained who was trained without describing how they were trained. None of the studies accounted for reorienting

coders to the tool if they had not recently participated in assessment for the study. Several studies measured assessment skill retention in analysis by measuring intrarater or test-retest reliability, in which a coder reassessed video of an assessment after one week, two months, or even a year since having first seen it. Studies measuring intrarater reliability included the PIPP (Slater et al., 2008), the N-PASS (Hummel, Lawlor & Weiss, 2009), and the ABC scale (Bellieni et al., 2005).

Translatability.

Many of the assessment tools validated in the selected studies were not conducted in the language or culture for which they were originally written. Three studies mentioned the process of translation, in which the tools were translated from English to another language and translated back again. The CHIPPS, NIPS, CRIES scales were translated from English to Thai (Suraseranivongse et al., 2006) and the PIPP scale was translated from English to both Swedish (Johansson & Kokinsky, 2009) and Norwegian (Vederhus et al., 2006). The correlation of the same scale in two different languages illustrates the translatability of the scale across different languages and cultures, regardless of both the cultural background of the infant, and the cultural influence of the person assessing for pain.

Nurse rating of clinical utility.

Four of the nineteen studies assessed clinical feasibility of tools as reported by nurse raters. Two of these studies validated the ABC scale (Bellieni et al., 2005; Bellieni et al., 2007a; Bellieni et al., 2007b; Suraseranivongse et al., 2006). In a 2007 study, ten nurse raters were asked to rank the three cry-related items in the ABC scale in relation to the preterm study sample as *unsatisfactory*, *good,* and *very good*. The raters described all three cues as *good* or *very good*. In a 2005 study with full term infants, nurse raters ranked the ABC in terms of time required,

simplicity, and utility as *good* out of a possible *unsatisfactory*, *improvable*, and *good*. In Suraseranivongse and colleagues' 2006 publication, nurses rated the CHIPPS, NIPS, and CRIES tools on a scale of one to ten regarding six components of clinical utility, including *ease of use*, *duration of rating, difficulty in assessment, help in decision to treat*, *feasible for clinical practice*, and *able to differentiate the severity of pain*. The NIPS scale, a unidimensional dichotomous tool, was deemed most clinically feasible in all areas, while CRIES, a multidimensional scale, was rated as the hardest to use, the most time consuming, the least helpful, and the least feasible. The CHIPPS scale was rated similarly to the NIPS scale, but 65% of nurses preferred NIPS to CRIES and CHIPPS. Nurses reported that the cue *posture of the trunk: rear up* was not useful in infants who were in a prone position (Suraseranivongse et al., 2006). In a 2007 study comparing the interrater reliability of NIPS and PIPP, the authors attributed the low agreement (Cohen's K=0.10) between a nurse at bedside and one via video to the inability of the bedside nurse to compile all the information for assessment into the two minute time period needed

Bedside versus video.

Within validation studies, infant pain tools are often used for assessment at bedside, or the infants are videotaped and rated for pain at a later time. Rating at bedside captures a realistic clinical scenario, but raters often cannot be blinded to the event, which can influence reliability and validity. Blinding raters and reducing bias can be made easier through coding by video. Nine of the studies used assessment at the bedside, validating the FLACC, CRIES, PIPP, NIPS, NFCS, N-PASS, PAT and COMFORT-B scales (Ahn & Jun, 2007; Ancora et al., 2009; Hummel et al., 2008; Johansson & Kokinsky, 2009; Jonsdottir & Kristjansdottir, 2005; Serpa et al., 2007; Spence et al., 2005; Taksande et al., 2005; Vederhus et al., 2006). One study was able to blind the raters while conducting the experiment with N-PASS at bedside (Hummel, Lawlor & Weiss,

2009). Six studies assessed with scales via video, helping validate the ABC scale, BIIP scale, and PIPP scales (Bellieni et al., 2005; Bellieni et al., 2007a; Eriksson et al., 2008; Holsti & Grunau, 2007b; Holsti et al., 2008; Slater et al., 2008). Two studies employed both video and bedside rating, with one study using video rating to determine construct validity and using bedside assessment to determine criterion validity for CRIES, CHIPPS and NIPS (Suraseranivongse et al., 2006), and another study using both video and bedside assessment to determine the interrater reliability of NIPS and PIPP (Bellieni et al., 2007b).

Seven of the thirteen tools reviewed contain cues that require the intervention or opinion of a nurse or caregiver to complete the assessment, including the PAT, EDIN, CRIES, COMFORT, COMFORT-B, FLACC, and N-PASS tools. The PAT, EDIN, and FLACC scale require input by the caregiving nurse; the PAT score includes the caregiving nurse's opinion of an infant's pain over an entire shift, the EDIN scale includes *quality of contact with nurses* as a cue and is used at the end of an eight hour observation period, and the FLACC includes consolability. Three tools require the assessment of reflex or tone; COMFORT and COMFORT-B require touching the infant to assess muscle tone, while the N-PASS assesses for grasp reflex. Three tools explicitly require extended observation; CRIES requires the infant to be observed the hour previous to assessment for *sleeplessness,* while the PAT's element of nurse opinion and the EDIN's *quality of contact with nurses* encompasses the infant's behavior over an entire shift.

The selected studies within this review encapsulate the current trends in infant pain tool research, specifically with the neonatal population. Through a variety of acutely ill and healthy preterm and full term neonates, researchers validated the selected pain scales for research and clinical settings and investigated the validity of cues like gestational age and cry. The aims of the studies, including the assessment of discriminant validity, focused on the current trend in

pain assessment toward the recognition of separate stress and pain states and the pursuit of a tool which can identify the difference in vulnerable preterm infants.

Chapter 5: Discussion

Within the spectrum of infant care, accurate assessment of pain and stress is key to reducing the negative impact that neonatal pain can have on infant development and care. Neonates residing in NICUs are exposed to a harsh environment and infant pain and stress assessment is not always consistent among hospital staff, which can lead to repeated neglect of an infant in pain (Barker & Rutter, 1995; Reyes, 2003). Assessment tools which can be consistently and practically applied to evaluate stress and pain in neonates are paramount to determining the need for treatment.

Within this review, studies comprised of full term and preterm samples of neonates were used to validate tools for the detection of acute pain, prolonged pain, general pain, and sedation through observational and quasi-experimental methods. To be considered a valid instrument, a tool must be tested for certain psychometrics, such as content, construct, and criterion validity, interrater and intrarater reliability, and internal consistency. Several tools were also tested for clinical utility by assessing nurse satisfaction. Out of the 13 tools included in this review, several emerged as the most valid and reliable for use in research and use in clinical practice. Specific tools were identified as the most valid for procedural pain, acute-prolonged pain, and sedation, which include the PIPP, NFCS, NIPS, N-PASS, and more.

Tools Suitable for Research

Instruments suitable for scientific inquiry should have specificity for pain and demonstrate validity and reliability. Tools preferable for research purposes often include a larger number of cues than do clinically useful tools; the additional information can more fully describe a range of infant behaviors and thus provide increased specificity. Tools measuring procedural pain are common in research, as procedures are more easily testable for the acute response they

invoke, rather than assessing for prolonged or chronic pain. The PIPP and the NFCS are two instruments that, through their design and validation, are well suited for pain assessment in research.

Premature Infant Pain Profile (PIPP).

The Premature Infant Pain Profile (PIPP) (Stevens et al., 1996) has been validated for procedural pain for all of the psychometrics previously mentioned. Many of the studies within this review use the PIPP scale as a gold standard to assess construct validity for other scales; the PIPP had acceptable agreement with both unidimensional and multidimensional scales including the ABC scale, FLACC scale, and N-PASS (Bellieni et al., 2007a; Ahn & Jun, 2007; Hummel et al., 2008). Researchers within this review also defined specific PIPP scores at which the infant could be considered in pain (Jonsdottir & Kristjansdottir, 2005). The PIPP scale has a large range, with seven indicators that can be scored from 0-3, which allows higher specificity in pain measurement. While the PIPP has excellent validity and reliability, some studies within the review reported poor clinical feasibility because of low interrater reliability (Ahn & Jun, 2007; Bellieni et al., 2007b). However, these results conflict with a previous validation study by the originators of the PIPP instrument where interrater reliability was 0.94-0.98 for video and bedside raters, respectively (Ballantyne, Stevens, McAllister, Dionne & Jack, 1999).

Additionally, the PIPP scale has demonstrated translatability to Norwegian and Swedish languages. Pain tools should be tested for translatability, not only so the tools are available in several different languages, but also so the tools can be considered reliable in different countries, regardless of the influence that culture may have on the assessment of pain.

Neonatal Facial Coding System (NFCS).

Another infant pain tool suitable for research in neonates is the NFCS (Grunau & Craig, 1987), a unidimensional tool that assesses ten facial actions in response to procedural pain. Although the tool only appeared in one study within this review, it is a very important tool in the spectrum of infant pain assessment, and deserves elaboration. Originally created based from research of full term infants, this tool has previously been validated for both the preterm and full term population and at the bedside. In a 2009 study assessing infants during immunizations, the NFCS was not able to discriminate between the pain reactions of infants who were rated as either stressed or without stress before receiving an injection (Kohut & Riddell, 2009). The ability of the NFCS to discriminate between pain preceded by stress and pain not preceded by stress has not been tested with the NFCS in neonates. However, the NFCS has shown discriminant validity between the constructs of non-pain and pain, through neonates undergoing either a heel rub or a heel lance (Grunau & Craig, 1987). This tool has been used repeatedly in research and is often the basis for studying facial reactivity in infants.

Certain cues within the NFCS may be less applicable to certain populations; the items *nasolabial furrow, open lips, stretched mouth, lip purse,* and *taut tongue* may be affected by intubation (Grunau, Oberlander, Holsti & Whitfield, 1998). However, a study within this review did use the tool with a sample that included ventilated infants (Serpa et al., 2007). Because very preterm infants may not have the ability to demonstrate the lower facial discriminant characteristics of lip purse and taut tongue, these indicators may be unreliable for this population (Haidet et al., 2009). The indicators *nasolabial furrow, brow bulge,* and *eye squeeze* have been proven to be as predictive of pain as the complete tool, so the NFCS is still reliable for these populations (Peters et al., 2003)

Tools Suitable for Clinical Use

Some tools are more suited for clinical use, based on the clarity of the items, the simplicity of the scoring system, the ability to collect information without disturbing the infants, and the speed of assessment.

ABC scale *(*Acuteness of the first cry, Burst rhythmicity, Constancy of cry).

The ABC scale (Bellieni et al., 2005) is a three-cue tool with a possible score from 0-6 that assesses cry based on the pitch of the first cry, the siren-like quality of the cry, and the duration of the cry. This tool was created for its clinical utility, as the assessor only needs to interpret auditory stimuli to determine an infant's score. The tool was tested with both full term and preterm samples in two separate studies within this review. Concurrent validity was greater with term infants, which may be a reflection of the decreased content validity of the ABC scale in the preterm population; preterm infants have a weaker cry response, and are more likely to not cry in response to pain in comparison to full term infants. Additionally, the tool is not validated for intubated patients because it cannot be administered on patients who cannot cry. The tool was tested for clinical feasibility in the two studies within the review, and nurses rated the tool as easy to use for both samples. Overall, the ABC scale is useful in a clinical setting as a fast way to assess pain in full term, healthy infants, but is less useful in preterm and/or intubated patients.

Neonatal Infant Pain Scale (NIPS).

The NIPS scale (Lawrence et al., 1993) is a six cue behavioral tool that measures five cues as 0 or 1 point and *cry* as 0, 1, or 2 points (Lawrence et al. 1993). The fact that a rater only chooses between two possible options for most of the items makes the tool easy to remember and report, but may lead to less specificity than a more detailed tool. It has previously been validated with preterm and full term infant samples. The inclusion of a *cry* and *respiration* cue makes the

tool less valid for intubated infants, although a study within the review used the tool to assess a sample that was 50% intubated. Within that study, nurses rated NIPS as the best tool for clinical use in comparison to CHIPPS and CRIES (Suraseranivongse et al., 2006), although it is not clear how easily and reliably the tool could have been applied to this intubated sample if cry and respiration could not be assessed as directed by the NIPS scale. Within that study, the tool was also translated accurately to Thai, which shows that it has translatability to other languages and cultures.

Behavioral Indicators of Infant Pain (BIIP).

The BIIP scale is a unidimensional tool validated for preterm infants that is comprised of eight different cues (Holsti & Grunau, 2007b). The tool includes five items from the NFCS, two hand movements, and a rating of behavioral state. Within the studies under review, the tool is tested for construct validity, criterion validity, discriminant validity, interrater reliability, and internal consistency. Raters using the BIIP assigned significantly different scores to infants undergoing stressful versus painful procedures. Although the researchers reported the BIIP as useful clinically, additional research is required to assess whether clinicians find the tool easy to use. The BIIP score can be used in intubated patients and with patients whose lower extremities are bundled.

Tools for Prolonged Pain and Sedation

Recently, researchers have developed more tools to test prolonged pain. Within this review, procedural pain was often assessed postoperatively with the EDIN, N-PASS, PAT, COMFORT, and COMFORT-B scales.

Neonatal Pain, Agitation, and Sedation Scale (N-PASS).

Among the tools used for prolonged and acute-prolonged pain, the N-PASS is a new multidimensional, five cue tool for preterm and full term infants (Hummel, et al., 2008). N-PASS is both a sedation tool and a pain tool, with scores of negative 10 to 0 representing sedation and 0 to positive 10 representing pain. The cues include crying/irritability, behavior/state, facial expression, extremities tone, and vital signs, which can be scored as -2, -1, 0, 1, or 2 points. An infant with a score of 0 is without pain and is not sedated. The tool has construct, discriminant, and criterion validity, and interrater and intrarater reliability, as shown in the studies within this review. N-PASS is validated for both acute-prolonged pain and procedural pain in the studies and has not been rated for clinical feasibility by nurses (Hummel et al., 2008; Hummel et al., 2009).

Échelle Douleur Inconfort Nouveau-né (EDIN).

The EDIN scale is also a new scale designed to assess prolonged pain. EDIN is a five item tool which rates cues like facial activity, body movement, and quality of sleep in addition to less traditional cues like, quality of contact with nurses, and consolability (Debillon et al., 2001). Cues are rated from 0-3 for a possible 0-15 point score. The EDIN scale was previously validated for content validity, construct validity, interrater reliability and intrarater reliability. Within the study in this review, researchers did not evaluate standard psychometrics, but investigated the legitimacy of adding additional points for infants with a low gestational age. The research validated adding 0-2 points to neonates' scores based on gestational age.

Pain Assessment Tool (PAT).

The PAT scale is a multidimensional tool for prolonged pain in preterm and full term patients (Hodgkinson et al., 1994). It was created in 1994 to be a reliable clinical tool for general

pain assessment and was not further validated for psychometrics until the 2005 study within this review which tested for interrater reliability and criterion validity. The tool has nine different behavioral and physiological items and one item titled *nurse perception of pain status* of the infant. This item's reliability may be affected by the experience of the nurse making the assessment; all of the nurses within the study had worked for a minimum of eight years in NICU nursing, and had used the PAT previously in clinical practice. The interrater reliability of the PAT cannot be generalized to a wide range of nurses. Overall, the PAT score needs to be tested with a wider range of raters for interrater reliability, intrarater reliability, and clinical feasibility to be a practical pain tool among for all nurses.

COMFORT -B (Behavioral) and COMFORT scales.

The COMFORT and COMFORT-B scales are used for intubated patients to monitor their level of sedation (Ambuel et al., 1992; van Dijk et al., 2000). The COMFORT-B was used in a study within this review as a unidimensional alternative to the COMFORT score, which assesses behavioral and physiological items. Both of the tools were originally validated for pediatric patients, and have been tested for fewer psychometrics for the neonatal population. Within the studies in this review, the COMFORT score was tested for criterion validity, sensitivity, specificity, and interrater reliability in premature ventilated neonates, whereas the COMFORT-B score has been tested for criterion validity and interrater reliability in the infant population. Both of these tools need more research concerning the neonatal population.

Tools for Postoperative Pain

Face, Legs, Activity, Cry, Consolability (FLACC).

The FLACC scale (Merkel et al., 1997) is a five-item unidimensional postoperative pain tool with a possible score of 0-10 that was created for clinical utility in preverbal patients.

Within the studies reviewed, the FLACC was validated for the neonatal population in regards to construct validity, criterion validity, and interrater reliability (Ahn & Jun, 2007). The FLACC assessed infants undergoing auditory stimuli, cluster care, and invasive procedures with significantly different scores. More research is needed to determine the clinical feasibility of the tool as reported by raters.

Crying, Requires oxygen, Increased vital signs, Expression, Sleeplessness (CRIES).

The CRIES scale (Krechel & Bildner, 1995) is a multidimensional postoperative pain tool created for neonates. The tool was previously validated for content, construct, and criterion validity and interrater reliability by the creators. Within the studies in this review, it was again tested for interrater reliability and criterion validity, but it was also tested for discriminant validity and nurse-rated feasibility. Suraseranivongse and colleagues (2006) tested the tool during a pain and non-pain setting, while Ahn and Jun (2007) tested the tool with three different levels of stimuli, as they did the FLACC scale. Due to the need for monitoring physiological parameters, nurses rated the CRIES scale as the least useful in a clinical setting in comparison to the CHIPPS and NIPS scales. Additionally, the CRIES scale was translatable to the Thai language and culture, and has translatability across cultures.

The Children's and Infant's Postoperative Pain Scale (CHIPPS).

The CHIPPS scale (Büttner & Finke, 2000) is a five item unidimensional postoperative pain tool that was initially created for infants and young children. Within this review, it is tested for sensitivity and specificity, criterion validity, interrater reliability and rated by nurses for feasibility. Nurses using the CHIPPS as it translated to Thai rated the tool as easy to use, but 65% of the raters preferred the NIPS scale and 20% preferred the CRIES scale to the CHIPPS

scale. Some nurses questioned the use of the *posture of the trunk* cue for use in prone children, as an infant who is prone cannot raise his or her trunk.

The Assessment of Stress

Nine studies within the review assessed infants at both a *pain* event and either a *non-pain* or *stress* event. Within the five studies that tested tools' ability to differentiate between a *pain* and a *non-pain* event, most pain events were defined as heel sticks or other painful, invasive procedures and corresponding non-pain events were defined as the preparatory events for the procedure. By separately assessing all preparatory handling surrounding the painful event, researchers ruled out the chance that a painful reaction could have been caused by anything besides the painful stimulus, and determined whether tools could detect any difference between an infant at baseline and an infant undergoing a non pain event. The ability of a pain tool to distinguish two constructs, in this case pain and --pain, is *discriminant validity*. When tested with the ABC scale (Bellieni, 2005; Bellieni et al., 2007a), BIIP scale (Holsti et al., 2008), CHIPPS, CRIES, NIPS (Suraseranivongse et al., 2006), N-PASS (Hummel et al., 2009), and PIPP scale (Eriksson et al., 2008), all found a significant difference between the two events. While non-pain is an issue, several studies investigated disruptive and possibly stressful procedures in comparison to pain.

In four studies within the sample, researchers tested tools' ability to differentiate between events that were defined as *painful* and events that were defined as *stressful* or *uncomfortable* (Ahn & Jun, 2007; Holsti et al., 2008; Jonsdottir & Kristjansdottir, 2005; Vederhus et al., 2006). These events were generally more disruptive than the *non-pain* events previously described and included diaper change and other elements of cluster care. The BIIP, FLACC, CRIES, and PIPP scales were assessed with these events, and the difference in scores between events for all tools

was significant. However, no studies reported an exact range in score which would specifically indicate stress.

Within the studies, researchers addressed the possibilities that stress induced before a painful procedure can increase the pain reaction to a later invasive event. For example, in the 2008 study about the BIIP and its ability to discriminate between pain and stress, Holsti and colleagues found that a painful event, defined as a heel lance, elicits an higher pain response when it is performed twenty minutes after a diaper change versus after a rest period of at least thirty minutes (Holsti et al., 2008). The stress of the NICU and cluster care may affect infant pain response, so simple procedures like a diaper change should be viewed as potentially detrimental if performed too frequently.

Another point that emerged from the research was the idea of what constitutes a painful event. In one study, researchers define endotracheal suctioning, a non-skin-breaking event as invasive and painful (Ahn & Jun, 2007). The FLACC, CRIES, and PIPP scores rated for this event support that definition, as suctioning scores were very similar to skin breaking procedures. Events which are known to cause discomfort in the adult population should also be expected to elicit a pain response in neonates.

Gestational Age as a Factor

Gestational age (GA) was researched in several studies within the review as an additional factor affecting reliable pain scores. Some tools employ gestational age as an item, adding additional points to the pain scores of infants born at an early gestational age. The results of the studies in this review both confirm and reject the idea of assigning additional points to infants born prematurely. In support of gestational age as a factor, the N-PASS and PIPP assign additional points to infants with lower GAs, due to previous research that suggests preterm

infants are less behaviorally competent in their repertoire of response to pain. Within this review, Ahn and Jun found that full term infants had higher scores than preterm infants undergoing the same painful and non-painful routine care procedures when assessed with the CRIES and FLACC, two tools that ignore gestational age. Preterm and full term infants assessed with PIPP during these procedures had similar scores (Ahn & Jun, 2007). Another study validated the addition of an extra point on the N-PASS for those infants below 30 weeks gestational age, as infants <30 weeks had lower pain scores than those >30 weeks (Hummel et al., 2009). Researchers studying the EDIN scale also found that neonates with lower gestational ages routinely had lower pain scores than full term infants (Ancora et al., 2009). These three studies confirm a difference between preterm and full term pain expression as measured by these tools, assuming that preterm and full term infants experience the same level of pain in response to identical stimuli.

Other studies had results that conflicted with the idea of adding extra points for lower gestational age; these studies either did not find a difference between the pain scores of preterm and full term infants or witnessed an increased reactivity to pain in the preterm infant. In Spence and colleagues' study validating the PAT scale, there was no significant difference between the scores of preterm and full term infants during the same painful event. However, the PAT scale includes a cue that uses the assessing nurse's opinion of whether or not the infant appears to be in pain. The nurses in this study had at least eight years of neonatal experience; their astute observation of preterm infants may have positively biased the average pain score for that age group (Spence et al., 2005). A study validating the PIPP scale found that preterm infants assessed with the PIPP had much higher pain scores than infants born at a later gestational age.

The authors reported that the sample sizes were too small to make any significant conclusions (Vederhus et al., 2006).

Overall, more research is necessary to add to the existing evidence supporting the addition of points to pain scores for lower gestational age. More studies stratifying by gestational age are needed to evaluate the subtle differences in behavioral repertoire by developmental stage.

Critiques of the Literature

Infant Age.

Several studies within this review did not clearly state elements of their samples' ages, whether it be postnatal age, gestational age at birth, or the range of the age of the sample. Consistent report of all aspects of preterm and full term infant age is necessary to evaluate prenatal and postnatal development, and contribute to the infant pain research concerning development.

Clinical Utility.

Several of the tools within this review were assessed for clinical feasibility through nurse report, such as the ABC scale, CHIPPS, NIPS, and CRIES scale. Still more tools claimed clinical feasibility without surveying the raters to determine whether they truly felt it was feasible. A tool may be considered useful due to the nature of its cues and the total number of items, but until it is tested for feasibility in the clinical setting, it cannot claim feasibility. The BIIP scale, FLACC scale, N-PASS, PAT, and PIPP scales claim clinical feasibility without reporting any nurse rating of the tools.

Criterion Validity.

Several studies within this review tested criterion validity with tools that may have been too divergent to produce favorable statistics. For example, in validating the unidimensional BIIP scale, comparison was drawn to the multidimensional NIPS scale, and the agreement found was lower than expected by the author at r=0.64 (Holsti & Grunau, 2007b). Some tools do not have a well documented gold standard and used expert opinion to establish criterion validity. In order to find the best criterion validity, tools should be compared to pre-existing tools that assess the same construct with a similar score range and the same type of cues, whether they are behavioral, physiological, or both.

Testing Cry Cues.

Several studies within this review validated tools that assess cry with populations that included intubated patients, such as NIPS (Serpa et al., 2007), PAT (Spence et al., 2005), CRIES, and CHIPPS (Suraseranivongse et al., 2006). In order to reliably use a tool, all cues must be applicable to the population being assessed. One study included modified versions of the FLACC and COMFORT-B scales to assess for *cry face* instead of *cry*. The N-PASS uses cry and irritability interchangeably in order to accommodate for intubated infants. Both of these approaches are appropriate.

Testing Reliability of Raters.

Within the nineteen studies under review, raters were used to assess infants with pain or stress tools to test for reliability. To ensure that results can be generalized to all potential raters, authors need to establish regular training methods for raters and report that training consistently. Most of the nineteen studies described the people who would assess infant pain or stress, including whether they assessed at bedside, by video, or both, and whether they were an

experienced or novice healthcare professional or research assistant. However, very few studies elaborated beyond that. In 2009, Haidet, Tate, Divirgilio, Kolanowski, and Happ outline specific ways to ensure rater reliability and retention, including the assessment of training, rater drift, and rater fatigue. Only four studies within this review described any sort of training process, and none of the studies described how the researcher would account for reteaching the raters in the event that an extended period of time passed between ratings. Haidet and colleagues emphasize that raters should be refreshed concerning the tools in order to avoid rater drift. Rater drift can be identified by testing intrarater reliability by requiring a rater to reassess a selection of taped events that he or she has previously coded. By comparing the scores that one rater assigns to an identical video on two separate occasions, the reliability of the instrument and of the training implemented can be assessed. Rater fatigue can also be assessed by requiring raters to rescore a selection of videos during the initial testing to ensure reliability throughout a coding session. While three of the studies under review reported intrarater reliability, none of the studies discussed rater fatigue. More emphasis needs to be placed on the reliability of the rater making assessments, not just the validity of the tool used.

Limitations of this Study

Many of the limitations of the studies are related to a lack of report; studies inconsistently reported gestational age at birth, postnatal age, and the rater training process. Additionally, those tools that were created for clinical utility often lacked any measure of nurse feedback; the tools were deemed feasible for practice without consulting the people who assessed with them. Some studies also used tools which assess cry in populations that, due to intubation, could not mount an audible cry response. The study samples also varied greatly in terms of location; there may be

cultural differences in the assessment and measurement of pain which were not taken into consideration.

Study Strengths

Strengths of this critical review of literature include the analysis of current trends within pain and stress assessment research, the presentation of current psychometric data for the most common and most recently researched assessment tools, and the thorough process used in searching the literature through multiple modalities. Also, the critique of the present tendency in research to underreport important elements of rater education and reliability is a necessary reminder to researchers and those utilizing these tools.

Implications for Nursing Practice

Within neonatal care, infant pain and stress assessment is key to delivering complete care; the healthcare provider that is able to interpret infant behavior and physiological status in the most definitive way possible can more accurately treat and care for the patient. Additionally, pain researchers with this ability can compile the most salient and correct information about their sample. When appropriately utilized, infant pain and stress tools can make a health professional's assessment more complete. The instruments evaluated in this review have a range of psychometric data and specific situations to which they are most appropriate; the organization of psychometric and evaluated feasibility of the scales within this review provides researchers and clinicians with a comprehensive background for choosing a scale that is best suited for their purposes. Also, healthcare professionals previously acquainted with these scales can also use the information presented in this review as an update, as the studies presented encompassed every recent validation article concerning these tools.

In addition to providing researchers and clinicians with information for choosing the most

appropriate pain tools for their use, the information in this review provides health care professionals with a synopsis of the items most used in pain and stress assessment and the cues associated with both states. Stress can be a catastrophic state for full term and preterm infants. Education about the importance of assessing an infant during both pain and non-pain situations may reinforce the need for a sensitive approach to neonatal care, especially for the preterm population.

Implications for Future Research

The studies within this review summarize the current research regarding infant pain and stress tools, and their analysis reveals both the direction of current research and the changes necessary to make that research more reliable. As current tools are validated for more populations and additional constructs, more emphasis should be placed on appropriately educating raters, using suitable samples for the tool being validated, reporting complete descriptive statistics, and testing for clinical feasibility.

For researchers, the information presented within this review may aid in the selection process for the most appropriate pain tool to use in assessment of a particular population. Also, the availability of assessment tools for both acute and prolonged pain creates more possibilities for an institution's standardized approach to infant pain and stress assessment.

Researchers are currently testing the discriminant validity of tools between painful and stressful events. As the definitions between stress and pain become more clear, careful and accurate assessment of stress and pain may better determine the state of an infant and accordingly, direct appropriate and successful care.

Reference List

Abdulkader, H., Freer, Y., Fleetwood-Walker, S., & McIntosh, N. (2008a). Bodily progression of motor responses to increasing mechanical force stimulation in the newborn infant and the effect of heel prick, *Neonatology, 94*(1), 38-44. doi:10.1159/000112948

Abdulkader, H., Freer, Y., Garry, E., Fleetwood-Walker, S., & McIntosh, N. (2008b). Prematurity and neonatal noxious events exert lasting effects on infant pain behavior. *Early Human Development, 84*(6), 351-355. doi:10.1016/j.earlhumdev.2007.09.018

Abedi, J. (1996). Interrater/test reliability system. *Multivariate Behavioral Research, 31*(4), 409-417. doi:10.1207/s15327906mbr3104_1

Ahn, Y. & Jun, Y. (2007). Measurement of pain-life response to various NICU stimulants for high-risk infants. *Early Human Development, 83*(4), 255-262. doi:10.1016/j.earlhumdev.2006.05.022

Als, H. (1982). Toward a synactive theory of development: Promise for the assessment and support of infant individuality. *Infant Mental Health Journal, 3*(4), 229-243. doi: 10.1002/1097-0355(198224)3:4<229::AID-IMHJ2280030405>3.0.CO;2-H

Als, H., Gilkerson, L., Duffy, F., McAnulty, G., Buehler, D., VandenBerg, K. … Jones, K. (2003). A three-center randomized controlled trial of individualized developmental care for very low-birth-weight preterm infants: Medical, neurodevelopmental, parent and care giving effects. *Journal of Developmental and Behavioral Pediatrics, 24*(6), 399-408. doi:10.1097/00004703-200312000-00001

Als, H., Lawhon, G., Duffy, F., McAnulty, G., Gibes-Grossman, R., & Blickman, J. (1994). Individualized developmental care for the very low-birth-weight preterm infant. *Journal of the American Medical Association. 272*(11), 853-858. doi:10.1001/jama.272.11.853

Ambuel, B., Hamlett, K., Marx, C., Blumer, J. (1992). Assessing distress in pediatric intensive care environments: the COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi:10.1093/jpepsy/17.1.95

Anand, K.J.S. (1998). Clinical importance of pain and stress in preterm neonates. *Biology of the Neonate, 73*(1), 1-9. doi:10.1159/000013953

Anand, K.J.S. (2007). Pharmacological approaches to the management of pain in the neonatal intensive care unit. *Journal of Perinatology, 27*(Suppl), S4–S11. doi:10.1038/sj.jp.7211712

Anand, K., Sippell, W., & Aynsley-Green, A. (1987). Randomised trial of fentanyl anaesthesia in preterm babies undergoing surgery: effects on the stress response. *The Lancet, 329*(8524), 62-66, 1987. doi:10.1016/S0140-6736(87)91907-6

Anand, K.J., Stevens, B.J., & McGrath, P.J. (2007). Pain in Neonates (3rd ed.). Philadelphia, PA: Elsevier Heath.

Ancora, G., Mastrocola, M., Bagnara, C., Zola, D., Pierantoni, L., Rossi, G.,…Faldella, G. (2009) Influence of gestational age on the EDIN score; an observational study. *Archives of Disease in Childhood: Fetal and Neonatal Edition, 94*, F35-F38. doi:10.1136/adc.2007.133777

Ballantyne M., Stevens B., McAlister M., Dionne K. & Jack A. (1999). Validation of the Premature Infant Pain Profile in the clinical setting. *Clinical Journal of Pain, 15* (29), 297–303. doi:10.1097/00002508-199912000-00006

Barker, D.P. & Rutter, N. (1995). Exposure to invasive procedures in neonatal intensive care unit admissions. *Archives of Disease in Childhood, 72*(1) F47-F48. doi:10.1136/fn.72.1.F47

Bauer, K., Ketteler, J., Hellwig, M., Laurenz, M., & Versmold. H. (2004). Oral glucose before

venepuncture relieves neonates of pain, but stress is still evidenced by increase in oxygen consumption, energy expenditure, and heart rate. *Pediatric Research, 55*(4), 695-700. doi:10.1203/01.PDR.0000113768.50419.CD

Becker, P., Grunwald, P., Moorman, J., & Stuhr, S. (1993). Effects of developmental care on behavioral organization in very-low-birth-weight-infants. *Nursing Research, 42*(4), 214-220. doi:10.1097/00006199-199307000-00005

Bellieni, C., Cordelli, D., Caliani, C., Palazzi, C., Franci, N., Perrone, S.,…Buonocore, G. (2007). Inter-observer reliability of two pain scales for newborns. *Early Human Development, 83*(8), 549-552. doi:10.1016/j.earlhumdev.2006.10.006

Bellieni, C., Bagnoli, F., Sisto, R., Neri, L., Cordelli, D., & Buonocore, G. (2005). Development and validation of the ABC pain scale for healthy full-term babies. *Acta Paediatrica, 94*(10), 1432-1436. doi:10.1080/08035250510039919

Bellieni, C., Maffei, M., Ancora, G., Cordelli, D., Mastrocola, M., Faldella, G.,…Buonocore, G. (2007). Is the ABC pain scale reliable for premature babies? *Acta Paediatrica, 96*(7), 1008-1010. doi:10.1111/j.1651-2227.2007.00355.x

Bellieni, C., Sisto, R., Cordelli, D. & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: an alarm threshold. *Pediatric Research, 55*(1), 142-146. doi:10.1203/01.PDR.0000099793.99608.CB

Blauer, T. & Gerstmann, D. (1998). A simultaneous comparison of three neonatal pain scales during common NICU procedures. *The Clinical Journal of Pain, 14*(1), 39-47. doi:10.1097/00002508-199803000-00006

Brazelton, T.B. (1973). *Neonatal Behavioral Assessment Scale: Vol. 50. Clinics in*

Developmental Medicine. London: Willian Heineman Medical Books. Philadelphia: J.B. Lippincott.

Büttner, W. & Finke W. (2000). Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: a comprehensive report on several consecutive studies. *Paediatric Anaesthesia, 10*(3), 303-18. doi:10.1046/j.1460-9592.2000.00530.x

Carbajal, R., Lenclen, R., Jugie, M., Paupe, A., Barton B., & Anand K. (2005). Morphine does not provide adequate analgesia for acute procedural pain among preterm neonates. *Pediatrics, 115*(6), 1494-1500. doi:10.1542/peds.2004-1425

Cornblath, M. & Ichord, R. (2000) Hypoglycemia in the neonate. *Seminars in Perinatology, 24*(2), 136-149. doi:10.1053/sp.2000.6364

Debillon, T., Zupan, V., Ravault, N., Magny, J., Dehan, M. & Abu-Saad, H. (2001). Development and initial validation of the EDIN scale, a new tool for assessing prolonged pain in preterm infants. Archives of Disease in Childhood. *Fetal and Neonatal Edition, 85*(1), F36-F41. doi:10.1136/fn.85.1.F36

Derbyshire, S.W.G., (1999). Locating the beginnings of pain. *Bioethics, 13*(1), 1-31. doi:10.1111/1467-8519.00129

DeVon, H., Block, M., Moyle-Wright, P., Ernst, D., Hayden, S., Lazzara, D.,…Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship, 39*(2), 155-164. doi:10.1111/j.1547-5069.2007.00161.x

Duhn, L.J. & Medves, J.M. (2004). A systematic integrative review of infant pain assessment tools. *Advances in Neonatal Care, 4*(3), 126-140. doi:10.1016/j.adnc.2004.04.005

Dunstan, P. (2009). Child sense: From birth to age 5, how to use the 5 senses to make sleeping,

eating, dressing and other everyday activities easier while strengthening your bond with your child. New York: Random House, Inc.

Eriksson, M., Storm, H., Fremming, A., & Schollin, J. (2008). Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. *Acta Paediatrica, 97*(1), 27-30. doi:10.1111/j.1651-2227.2007.00586.x

Fitzgerald, M., Millard, M., & MacIntosh, N. (1988). Hyperalgesia in premature infants, *The Lancet, 331*(8580), 292. doi:10.1016/S0140-6736(88)90365-0

Franck, L. & Miaskowski, C. (1997). Measurement of neonatal responses to painful stimuli: A research review. *Journal of Pain and Symptom Management, 14*(6), 343-378. doi:10.1016/S0885-3924(97)00222-4

Gibbins, S., Stevens, B., McGrath, P., Yamada, J., Beyene, J., Breau, L.,…Ohlsson, A. (2008). Comparison of pain responses in infants of different gestational ages. *Neonatology, 93*(1), 10-18. doi:10.1159/000105520

Grunau, R.V. & Craig, K.D. (1987). Pain expression in neonates: facial action and cry. *Pain, 28*(3) 395-410. doi:10.1016/0304-3959(87)90073-X

Grunau, R., Holsti, L., Whitfield, M., & Ling, E. (1999) Are twitches, startles, and body movements pain indicators in extremely low birth weight infants? *The Clinical Journal of Pain, 16*(1), 37-45. doi:10.1097/00002508-200003000-00007

Grunau, R.V., Johnston, C.C., & Craig, K.D. (1990). Neonatal facial and cry responses to invasive and non-invasive procedures. Pain, 42(3), 295-305. doi:10.1016/0304-3959(90)91142-6

Grunau, R., Oberlander, T., Holsti, L., & Whitfield, M. (1998). Bedside application of the

Neonatal Facial Coding System in pain assessment of premature neonates. *Pain, 76*(3), 277-86. doi:10.1016/S0304-3959(98)00046-3

Grunau R.E., Oberlander T.F., Whitfield M.F., Fitzgerald, C., Morison, S., & Saul, J. (2001). Pain reactivity in former extremely low birth weight infants at corrected age 8 months compared with term born controls. *Infant Behavior and Development , 24*(1), 41–45. doi:10.1016/S0163-6383(01)00065-0

Haidet, K.K. (2005). *Biobehavioral responses to caregiving in very low birth weight preterm infants* (Doctoral dissertation, the Pennsylvania State University). Retrieved from libraries.psu.edu.

Haidet, K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A., & Happ, M. (2009). Methods to improve reliability of video-recorded behavioral data. *Research in Nursing and Health, 32*(4), 465-74. doi:10.1002/nur.20334

Harmon, S.L., & McManus, B.M. (2008). Developmentally supportive care. In J.P. Cloherty, E.C. Eichenwald & A.R. Stark (Eds), Manual of Neonatal Care (pp. 155) (5^{th} ed.). Philadelphia, PA: Lippincott, William & Wilkins.

Hellerud, B. & Storm, H. (2002). Skin conductance and behaviour during sensory stimulation of preterm and term infants. *Early Human Development, 70*(1-2), 35-46. doi:10.1016/S0378-3782(02)00070-1

Herrington, C.J., Olomu, I.N., & Geller, S.M. (2004). Salivary cortisol as indicators of pain in preterm infants: a pilot study. *Clinical Nursing Research, 13*(1), 53-68. doi:10.1177/1054773803259665

Hohmeister, J., Demirakça, S., Zohsel, K., Flor, H., & Hermann, C. (2009). Responses to pain in school-aged children with experience in a neonatal intensive care unit: Cognitive aspects

and maternal influences. *European Journal of Pain, 13*(1), 94-101. \ doi:10.1016/j.ejpain.2008.03.004

Holsti, L. & Grunau, R. (2007a). Extremity movements help occupational therapists identify stress responses in preterm infants in the neonatal intensive care unit: a systematic review. *Canadian Journal of Occupational Therapy, 74*(3), 183-194. Retreived from ProQuest Nursing & Allied Health Source in Oct 2008. http://ezaccess.libraries.psu.edu/login?url=http://medjournal.hmc.psu.edu:2085/pqdweb?did=1288084021&sid=2&Fmt=2&clientId=9874&RQT=309&VName=PQD

Holsti, L. & Grunau, R. (2007b). Initial validation of the Behavioral Indicators of Infant Pain (BIIP). *Pain, 132*(3), 264-272. doi:10.1016/j.pain.2007.01.033

Holsti, L., Grunau, R., Oberlander, T., & Osiovich, H. (2008). Is it painful or not? Discriminant validity of the Behavioral Indicators of Infant Pain (BIIP) scale. *Clinical Journal of Pain, 24*(1), 83-88. doi:10.1097/AJP.0b013e318158c5e5

Holsti, L., Grunau, R, Oberlander, T., & Whitefield, M. (2004). Specific newborn individualized developmental care and assessment program movements are associated with acute pain in preterm infants in the neonatal intensive care unit. *Pediatrics, 114*(1), 65-72. doi:10.1542/peds.114.1.65

Holsti, L., Grunau, R., Oberlander, T., Whitfield, M., & Weinberg, J. (2005). Body movements: An important additional factor in discriminating pain from stress in preterm infants. *Clinical Journal of Pain, 21*(6), 491-498. doi:10.1097/01.ajp.0000146163.30776.44

Hummel, P., Lawlor-Klean, P., & Weiss, M. (2009). Validity and reliability of the N-PASS assessment tool with acute pain. Journal of Perinatology, Retrieved from http://www.nature.com/jp/journal/vaop/ncurrent/abs/jp2009185a.html.

doi:10.1038/jp.2009.185

Hummel, P., Puchalski, M., Creech, S., & Weiss, M. (2008). Clinical reliability and validity of the N-PASS: Neonatal pain, agitation and sedation scale with prolonged pain. *Journal of Perinatology, 28*(1), 55-60. doi:10.1038/sj.jp.7211861

Jain, S., Kumar., P., & McMillan, D. (2006). Prior leg massage decreases pain response to heel stick in preterm babies. *Journal of Paediatrics and Child Health. 42*(9), 505-508. doi:10.1111/j.1440-1754.2006.00912.x

Jansen, M.P.M. (2008). *Managing Pain in the Older Adult*. New York: Springer Publishing Company.

Johansson, M. & Kokinsky, E. (2009). The COMFORT behavioral scale and the modified FLACC scale in pediatric intensive care. *Nursing in Critical Care, 14*(3) 122-130. doi:10.1111/j.1478-5153.2009.00323.x

Johnston, C. Stevens, B., Craig, K., & Grunau, R. (1993). Developmental changes in pain expression in premature, full-term, two- and four-month-old infants. *Pain, 52*(2), 201–208. doi: 10.1016/0304-3959(93)90132-9

Jonsdottir, R. & Kristjansdottir, G. (2005). The sensitivity of the premature infant pain profile – PIPP to measure pain in hospitalized neonates. *Journal of Evaluation in Clinical Practice, 11*(6), 598-605. doi:10.1111/j.1365-2753.2005.00603.x

Kahn, A., Groswasser, J., Franco, P., Scaillet, S., Sawaguchi, T., Kelmanson, I., & Dan B. (2003). Sudden infant deaths: stress, arousal and SIDS. *Early Human Development,* 75 Supplement, S147-166. doi:10.1016/j.earlhumdev.2003.08.018

Krechel, S. & Bildner, J. (1995). CRIES: a new neonatal postoperative pain measurement score;

Initial setting of validity and reliability. *Paediatric Anaesthesia, 5*(1), 53-61. doi:10.1111/j.1460-9592.1995.tb00242.x

Kohut, S.A., & 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. doi:10.1016/j.jpain.2008.08.010

Lawrence, J., Alcock, D., McGrath, P., Kay, J., MacMurray S., & Dulberg, C. (1993). The development of a tool to assess neonatal pain. *Neonatal Network, 12*(6), 59-66.

Luhr, V.T., Zeskind, P.S., Ofenstein, J.P., Cepeda, E.C., Warrier, I. & Aranda, J.V. (2007). Neonatal Facial Coding System scores and spectral characteristics of infant crying during newborn circumcision. Clinical Journal of Pain, 23(5), 417-424. doi:10.1097/AJP.0b013e31805476f2

Lucas-Thompson, R., Townsend, E., Gunnar, M., Georgieff, M., Guiang, S., Ciffuentes, R.,… Davis, E. (2008). Developmental changes in the responses of preterm infants to a painful stressor. *Infant Behavior & Development, 31*(4), 614–623. doi:10.1016/j.infbeh.2008.07.004

Ludington-Hoe, S., Hosseini, R., & Torowicz, D. (2005). Skin to skin contact (Kangaroo Care) analgesia for preterm infant heel stick. *AACN Clinical Issues, 16*(3), 373-387. doi:10.1097/00044067-200507000-00010

MacGregor R., Evans, D., Sugden, D., Gaussen, T., & Levene, M. (1998). Outcome at 5–6 years of prematurely born children who received morphine as neonates. *Archives of Disease in Childhood; Fetal and Neonatal Edition, 79*(1), F40–F43. doi:10.1136/fn.79.1.F40

Manfredi, C., Bocchi, L., Orlandi, S., Spaccaterra, L., & Donzelli, G. (2009). High-resolution cry analysis in preterm newborn infants. *Medical Engineering & Physics, 31*(5), 528-532.

doi:10.1016/j.medengphy.2008.10.003

McCullough, S., Halton, T., Mowbray, D., & MacFarlane, P. (2008). Lingual sucrose reduces the pain response to nasogastric tube insertion: a randomised clinical trial. *Archives of Disease in Childhood; Fetal and Neonatal Edition, 93*(2), 100-103. doi:10.1136/adc.2006.110338

McDowell, I. & Newell, C. (1996). Measuring health; a guide to rating scales and questionnaires. New York: Oxford University Press.

McEwen, B.S. (1998). Protective and damaging effects of stress mediators. *The New England Journal of Medicine*, 338(3), 171-179. doi:10.1056/NEJM199801153380307

McEwen, B.S. (2005). Stressed or stressed out: What is the difference? *Journal of Psychiatry and Neuroscience, 30*(5), 315-318. Retreived from Canadian Medical Association. http://www.cma.ca/multimedia/staticContent/HTML/N0/l2/jpn/vol-30/issue-5/pdf/pg315.pdf

Merkel, S., Voepel-Lewis, T., Shayevitz, J., & Malviya, S. (1997). The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing, 23*(3), 293-297.

Oberlander, T., Gruanu, R., Whitfield, M., Fitzgerald, C., Pitfield, S., & Saul, J. (2000). Biobehavioral pain responses in former extremely low birth weight infants at four months' corrected age. *Pediatrics,105*(1), e6. doi:10.1542/peds.105.1.e6

Pasero, C. (2002). Pain control: Pain assessment in infants and young children: neonates. *The American Journal of Nursing, 102*(8), 61-65. Retreived from Lippincott, Williams & Wilkins. http://www.jstor.org/stable/3523036.

Peng, N., Bachman, J., Jenkin, R., Chen, C., Chang, Y., Chang,, & Wang, T. (2009).

Relationships between environmental stressors and stress biobehavioral responses of preterm infants in NICU. *The Journal of Perinatal and Neonatal Nursing, 23*(4), 363-371. doi:10.1097/JPN.0b013e3181bdd3fd

Pereira, A., Guinsburg, R., de Almeida, M., Monteiro, A., dos Santos, A., & Kopelman, B. (1999). Validity of behavioral and physiologic parameters for acute pain assessment of term newborn infants. *São Paulo Medical Journal, 117*(2), 72-80. doi:10.1590/S1516-31801999000200005

Peters J., Koot, H., Grunau, R., de Boer, J., van Druenen, M.J. & Tibboel, D. (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. doi:10.1097/00002508-200311000-00003

Peters, J., Schouw, R., Anand, K., van Dijk, M., Duivenvoorden, H., & Tibboel, H., (2005). Does neonatal surgery lead to increased pain sensitivity in later childhood? *Pain, 114*(3), 444-454. doi:10.1016/j.pain.2005.01.014

Pressler, J.L., Hepworth, J.T., Helm, J.M., & Wells, N.L. (2001). Behaviors of very preterm neonates as documented using NIDCAP observations. *Neonatal Network: NN, 20*(8), 15-24.

Ranger, M., Johnston, C., & Anand, K. (2007). Current controversies regarding pain assessment in neonates. *Seminars in Perinatology, 31*(5), 283-288. doi:10.1053/j.semperi.2007.07.003

Reyes, S. (2003). Nursing Assessment of Infant Pain. The Journal of Perinatal & Neonatal Nursing, 17(4), 291-303. Retrieved from http://journals.lww.com/jpnnjournal/Abstract/2003/10000/Nursing_Assessment_of_Infant_Pain.8.aspx

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature, 138*, 32. doi: 10.1038/138032a0

Serpa, A., Guinsburg, R., Balda, R., dos Santos, A., Areco, K., & Peres, C. (2007). doi: Multidimensional pain assessment of preterm newborns at the 1st, 3rd and 7th days of life. *Sao Paulo Medical Journal, 125*(1), 29-33. doi:10.1590/S1516-31802007000100006

Shah, P., Aliwalas, L., & Shah, V., (2007). Breastfeeding or breastmilk to alleviate procedural pain in neonates: a systematic review. *Breastfeeding Medicine, 2*(2), 74-82. doi:10.1089/bfm.2006.0031

Slater, R., Canterella, A., Franck, L., Meek, J., & Fitzgerald, M. (2008). How well do clinical pain assessment tools reflect pain in infants? *Plos Medicine, 5*(6), e129. doi:10.1371/journal.pmed.0050129

Spence, K., Gillies, D., Harrison, D., Johnston, L., & Nagy, S. (2005). A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *Journal of Obstetric, Gynecologic, and Neonatal Nursing, 34*(1), 80-86. doi:10.1177/0884217504272810

Sredl, D. (2003). Myths and facts about pain in neonates. *Neonatal Network, 22*(6) 69-71.

Streiner, D.L. & Norman, G.R. (1994). *Health measurement scales; A practical guide to their development and use* (2nd ed.). New York: Oxford University Press.

Stevens, B., Johnston, C., Franck, L., Petryson, P., Jack, A., & Foster, G. (1999). The efficacy of developmentally sensitive interventions and sucrose for relieving procedural pain in very low birth weight neonates. *Nursing Research, 48*(1) 35-43. doi:10.1097/00006199-199901000-00006

Stevens, B., Johnston, C. & Horton, L. (1994). Factors that influence the behavioral pain

responses of premature infants. *Pain, 59*(1)101-109. doi:10.1016/0304-3959(94)90053-1

Stevens, B., Johnston, C., Petryshen, P., Taddio, A. (1996). Premature infant pain profile; development and initial validation. *Clinical Journal of Pain, 12*(1) 13-22. doi:10.1097/00002508-199603000-00004

Stevens, B., McGrath, P., Yamada, J., Gibbins, S., Beyene, J., Breau, L.,…Ohlsson, A. (2006). Identification of pain indicators for infants at risk for neurological impairment: A Delphi consensus study. *Biomedical Central Pediatrics, 6*(1), 1-9. doi:10.1186/1471-2431-6-1

Suraseranivongse, S., Kaosaard, R., Intakong, P., Pornsiriprasert, S., Karnchana, Y., Kaopinpruck, J., & Sangjeen, K. (2006). A comparison of postoperative pain scales in neonates. *British Journal of Anesthesia, 97*(4), 540-544. doi:10.1093/bja/ael184

Taksande, A., Vilhekar, K., Jain, M., Chitre, D. (2005). Pain response of neonates to venipuncture. *Indian Journal of Pediatrics, 72*(9), 751-753. doi:10.1007/BF02734146

Tinsley, H.E. & Weiss, D.J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*(4), 358-376. doi:10.1037/h0076640

van Dijk, M., de Boer, J., Koot, H., Tibboel, D., Passchier, J., & Duivenvoorden, H. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain, 84*(2-3), 367–377. doi:10.1016/S0304-3959(99)00239-0

Vederhus, B.J., Eide, G.E., Natvig, G.K. (2006). Psychometric testing of a Norwegian version of the Premature Infant Pain Profile: an acute pain assessment tool. A clinical validation study. *International Journal of Nursing Practice, 12*(6), 334-344. doi:10.1111/j.1440-172X.2006.00592.x

Wielenga, J., De Vos, R., de Leeuw, R., & De Haan, RJ. (2004). COMFORT scale: a reliable

and valid method to measure the amount of stress of ventilated preterm infants. *Neonatal Network: NN, 23*(2), 39-44. Retreived from George T. Harrell Library Journal List. http://medjournal.hmc.psu.edu:2290/content/121252/?sortorder=asc&Copyright=2004.

Williams, A.L., Khattak, A.Z., Garza, C.N., & Lasky, R.E. (2009). The behavioral pain response to heel stick in preterm neonates studied longitudinally: Description, development, determinants, and components. *Early Human Development, 85*(6), 369-374. doi:10.1016/j.earlhumdev.2009.01.001

ACADEMIC VITA of Molly K. McInnis

Molly K. McInnis
1506 Berwick Bend
Zachary, LA 70791
mollykathleenmcinnis@gmail.com

Education:

Bachelor of Science Degree in Nursing, Penn State University, Spring 2005
Bachelor of Arts Degree in Visual Arts, Penn State University, Spring 2005
Honors in Nursing
Thesis Title: Infant Pain and Stress Assessment; An Integrated Review
Thesis Supervisor: Dr. Kim Kopenhaver Haidet
Honors in Visual Arts
Thesis Title: We Were Soldiers; A Visual Biography
Thesis Supervisor: Helen O'Leary

Related experience:

PRN Nurse Extern at Hershey Medical Center
Supervisor: Rita Barry
November 2009-May 2010

Volunteer at Baton Rouge Medical Center
Supervisor: Kathleen Miller
May 2009-July 2009

Solo Gallery Exhibits for We Were Soldiers
We Were Soldiers" Patterson Gallery, University Park, PA, March 2009
"Art on the Move" Warnock Commons, University Park PA, January- March 2009
"Art on the Move" Old Main, University Park, PA, October 2008 – January 2009
"Art on the Move" Lewis Katz Building, University Park, PA, May-October 2009

Awards/Honors:

Deans List
Summer Discovery Research Grant
Member of Phi Beta Kappa
Member of Sigma Theta Tau, an International Nursing Honor Society
Nominated for the Arts and Architecture Golumbic Award, March 2009
2nd place in Undergraduate Research Exhibition, Arts and Humanities category

Activities:

Trash to Treasure Winter Sale Co-Coordinator, Spring 2009
Member of Student United Way, 2007-2009
Artistic Coordinator for the Tunnel of Oppression, Spring 2008
Student Nurses Association of Pennsylvania (SNAP)