



OPEN ACCESS

EDITED BY

Enrico Cocchi,
University of Bologna, Italy

REVIEWED BY

Mariam John Amin Ibrahim,
Ain Shams University, Egypt
Guzide Ugucu,
Mersin University, Türkiye
Zi Zeng,
Henan Vocational College of Nursing, China

*CORRESPONDENCE

Zehra Kan Öntürk
✉ zehra.kan@acibadem.edu.tr

RECEIVED 02 June 2025

REVISED 14 December 2025

ACCEPTED 18 December 2025

PUBLISHED 12 January 2026

CITATION

Albayrak S, Kan Öntürk Z, Şen E and Yayla M (2026) Neonatal intensive care nurses' assessment of preterm infants' pain and sedation: inter-rater reliability of the neonatal pain, agitation, and sedation scale. *Front. Pediatr.* 13:1639511. doi: 10.3389/fped.2025.1639511

COPYRIGHT

© 2026 Albayrak, Kan Öntürk, Şen and Yayla. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Neonatal intensive care nurses' assessment of preterm infants' pain and sedation: inter-rater reliability of the neonatal pain, agitation, and sedation scale

Selvinaz Albayrak¹ , Zehra Kan Öntürk^{2*} , Elif Şen³ and Melike Yayla³

¹Department of Nursing, Faculty of Health Sciences, Istinye University, Istanbul, Türkiye, ²Department of Nursing, Faculty of Health Sciences, Acibadem Mehmet Ali Aydınlar University, Istanbul, Türkiye, ³Neonatal Intensive Care Unit, Acibadem Health Group Altunizade Hospital, Istanbul, Türkiye

Background: Timely and accurate assessment of pain and sedation in newborns is essential for effective management. Therefore, neonatal pain and sedation assessment remains a key global issue in neonatal intensive care unit (NICU) nursing practice. This study examined the inter-rater reliability of Neonatal Pain Agitation, and Sedation Scale (N-PASS) scores among NICU patients.

Methods: This prospective observational study assessed agreement among 19 NICU nurses and two independent researchers who completed 190 observations from 82 preterm infants. Each evaluator rated N-PASS independently and blindly. Agreement among three raters—a nurse and two researchers—were analyzed using the intraclass correlation (ICC) and the Fleiss kappa test.

Results: Agreement levels varied across N-PASS subscales. The ICC and kappa values indicated moderate-to-good reliability for the pain/agitation subscale, whereas the ICC values for the sedation subscale indicated excellent or moderate reliability. Nurses assigned higher mean pain/agitation scores than researchers.

Conclusions: NICU nurses must improve their N-PASS assessment skills for both pain and sedation. NICU nurse managers should prioritize improving these competencies to improve pain experiences and ensure adequate sedation, given their significant impact on short- and long-term outcomes in preterm infants.

KEYWORDS

inter-rater reliability, neonatal intensive care units, neonatal nursing, neonatal pain assessment, pain management

1 Introduction

Pain is common among infants in neonatal intensive care units (NICUs), where many undergo repeated diagnostic and therapeutic procedures during hospitalization (1, 2). Because frequent painful exposures and early-life stress may negatively affect physiologic stability and neurodevelopment in preterm infants, accurate recognition and management of pain remains a critical clinical priority in NICUs (3–10).

Although the importance of neonatal pain assessment is well recognized, practices across NICUs remain inconsistent (1, 11, 12). Reliable evaluation of pain, agitation, and sedation is necessary to prevent both undertreatment and unnecessary exposure to analgesic or sedative medications (13). Pain assessment is challenging due to infants' inability to communicate verbally, gestational age-related differences in behavioral cues, and variability in clinician training and experience (14–16). Furthermore, physiologic and behavioral stress responses often resemble pain-related cues, complicating differentiation in clinical settings (14). This overlap can lead to misclassification, causing either insufficient analgesia or unwarranted pharmacologic treatment (17).

Consistent use of structured assessment tools is essential to support accurate clinical decision-making (18). Nurses play a central role in applying these tools in the NICU; however, studies have shown variability in assessment accuracy (19). Several pain assessment scales exist for neonatal populations (14, 20, 21). Among these, the Neonatal Pain, Agitation, and Sedation Scale (N-PASS) is widely used because it combines behavioral and physiologic indicators and is applicable across a broad gestational age range for assessing both pain and sedation (22–24). However, interpretation of these indicators in preterm and critically ill infants may differ among assessors, making inter-rater reliability an important component of the tool's clinical validity (14). Only a limited number of studies have evaluated the inter-rater reliability of the N-PASS in real-time clinical settings (22, 23, 25).

Therefore, this study aimed to evaluate the inter-rater reliability of the N-PASS in a real clinical setting. Considering that consistency among assessors is essential for accurate pain evaluation, this study specifically examined the consistency of NICU nurses' pain and sedation assessments using the N-PASS in preterm infants.

2 Methods

2.1 Study design, setting and participants

This prospective, blinded observational study examined agreement of N-PASS scores among NICU nurses. The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (26) and was registered in the Clinical Trial Registry of the U.S. National Institutes of Health (<https://clinicaltrials.gov/>; identifier: NCT06885437). All procedures adhered to the Declaration of Helsinki and were approved by the "Acibadem Mehmet Aydınlar University Medical Research Review Board (ATADEK (Decision no: 2022-20/25, December 30, 2022))."

Data collection took place between January and March 2023 in a 28-bed NICU staffed by 26 nurses working two 12-h shifts daily. Except for the charge nurse, all nurses rotated through both day and night shifts based on the weekly staffing schedule. Inclusion criteria required at least 6 months of NICU experience and willing to participate; nurses with <6 of experience were

excluded. Before enrollment, nurses received information about the study purpose, procedures, voluntary participation, and the right to withdraw at any time. Both verbal and written informed consent were obtained. Two nurses (7.7%) declined participation, and five (19.2%) did not meet eligibility requirements, resulting in a final sample of 19 nurses (73.1%). Data were collected during day-shift assignments.

Following recommendations by Koo and Li (2016), which emphasize obtaining a heterogeneous sample of at least 30 participants and including at least three raters in reliability studies, this study included 190 observations from 82 preterm infants younger than 37 gestational weeks. Repeated observations for the same infant were allowed only when separated by at least 24 h. Inclusion criteria for infants were all preterm infants admitted during the study period; exclusion criteria were major congenital anomalies, severe neurologic disorders, or paralysis conditions that could affect accurate pain assessment. Pain and sedation assessments from 190 observations were completed simultaneously by three raters—two researchers and one nurse. Each of the 19 nurses assessed 10 infants, yielding a total of 190 infant assessments (19 nurses × 10 infants). Thus, two researchers and one nurse evaluated each infant together, consistent with rater-based clinical assessment methods described on the literature (27).

The N-PASS scale has been routinely used in the NICU since 2020 to assess pain and agitation 6 times daily for all infants, and sedation for those receiving sedative medications. At implementation, nurses completed structured training covering key principles of neonatal pain assessment and practical scoring procedures. Newly hired nurses received the same orientation during onboarding. Therefore, all raters in this study were familiar with the N-PASS and competent in its routine clinical application.

2.2 Instruments

2.2.1 Nurse demographics and training for pain assessment and management

This section included six research-developed questions covering nurses' age, sex, educational status, total professional experience, NICU experience, and prior training in neonatal pain and sedation assessment and management.

2.2.2 Infant demographics

This section included 10 research-developed questions regarding each infant's sex, gestational age at birth, birth weight, current weight, diagnosis, type of ventilation support, sedation interventions, and pain relief interventions.

2.2.3 Neonatal pain agitation and sedation scale (N-PASS)

The N-PASS, developed by Hummel et al. (2008), consists of two subscales, with pain/agitation and sedation scores calculated separately (22, 24, 28, 29). Each subscale includes four behavioral items—crying or irritability, behavior state, facial

expression, and extremities tone—and one vital signs item assessing changes in heart rate, respiratory rate, blood pressure, and oxygen saturation (25). The N-PASS is used for premature infants, for assessing acute and chronic pain, and for infants receiving mechanical ventilation. It has been validated for infants from 23 gestational weeks to 36 months, making it suitable for preterm infants in this study, including those born at 24 weeks. Pain and agitation items are scored as 0, 1, or 2 (22, 25). “The maximum score is 11 for preterm infants younger than 30 weeks and 10 for those older than 30 weeks. Pain levels are categorized as no pain (0–3), mild pain (4–7), and severe pain (8–11)” (22, 30). Sedation scoring uses the same behavioral and physiologic categories, scored as 0, –1, or –2. The maximum is –10, with 0 indicating no sedation. Score from –1 to –5 reflect light sedation, and scores from –5 to –10 represent deep sedation (25). Cronbach’s alpha coefficients for the original pain and sedation subscales were 0.82 and 0.87, respectively (22, 25). The Turkish adaptation by Açıkgöz et al. (2017) reported Cronbach’s alpha values were 0.83 and 0.77 for the sedation subscale (31). In this study, Cronbach’s alpha values were 0.89 for the pain subscale and 0.95 for the sedation subscale. Other studies using the Neonatal Pain, Agitation, and Sedation Scale reported Cronbach’s alpha values of 0.93, 0.84, and 0.85 for the pain subscale (32–34) and 0.88 and 0.73 for the sedation subscale (28, 34). Overall, the N-PASS scale is a reliable tool for assessing acute pain in both mechanically and nonmechanically ventilated neonates, evaluating long-term pain in mechanically ventilated or postoperative infants, and determining sedation levels in mechanically ventilated or postoperative infants (29).

2.3 Data collection

Data collection was performed by two researchers and NICU nurses. Both researchers held Bachelor of Science in Nursing degrees, had 5 years of NICU experience, and were certified NICU nurses. One researcher served as a charge nurse and the other as a clinical nurse educator.

Each observation involved two fixed researchers and one bedside nurse. The two researchers evaluated all infants throughout the study, whereas the bedside nurse rotated according to shift assignments. Thus, each assessment was independently completed by a three-rater team: two consistent researchers and one rotating bedside nurse.

After identifying infants who met the inclusion criteria, the N-PASS scores for pain, agitation, and sedation were independently and blindly assessed by the three observers. The two researchers completed 190 assessments, and each of the 19 nurses scored 10 preterm infants. Pain and agitation assessments were performed after routine nursing care (monitoring vital signs, feeding support, positioning and comfort care, hygiene, and skin care) in 190 observations from 82 preterm infants. Sedation assessments were conducted after routine care in 46 preterm infants receiving sedatives. All assessments were completed simultaneously at the bedside through direct observation.

Each observer was given one minute to independently record the total Neonatal Pain, Agitation, and Sedation Scale score.

Demographic and clinical data for both nurses and infants were also documented. Blinding procedures were strictly followed. Each observer independently evaluated the same infant without access to the other raters’ scores. Observers recorded their assessments on separate data collection forms at the bedside without any verbal or visual communication about their evaluations. Nurses and researchers remained blinded to one another’s ratings and submitted their completed forms individually to the third researcher at the end of data collection.

2.4 Data analysis

There were no missing data. A total of 190 observations for pain and agitation and 46 observations for sedation were analyzed using Statistical Package for the Social Sciences version 26.0 (IBM, Armonk, NY, USA). Descriptive statistics were used to summarize the demographic characteristics of nurses and infants, as well as the N-PASS scores. Skewness and kurtosis values were examined to assess the normality of summed pain, agitation, and sedation scores. Inter-rater reliability was evaluated by comparing the three raters’ N-PASS pain and sedation scores. Intraclass correlation coefficient (ICC) values were calculated for both single and average measures. A two-way random-effects model was selected because it is appropriate for rater-based clinical assessment methods. Absolute agreement was used to determine whether different raters assigned identical values. ICC values were interpreted as follows: <0.50, poor reliability; 0.50–0.75, moderate reliability; 0.75–0.90, good reliability; and greater than 0.90, excellent reliability (27). The paired t-test was used to compare pain/agitation scores across the three raters. The Wilcoxon signed-rank test was used to compare sedation scores because of the smaller sample size.

N-PASS pain scores were categorized as no pain (0–3), mild pain (4–7), and severe pain (8–11). Sedation scores were categorized as “no sedation (0), light sedation (–1 to –5), and deep sedation (–6 to –10). Fleiss kappa statistics with absolute agreement were used to assess the inter-rater reliability of the three observers for these categorical variables. Kappa (κ) values were interpreted as follows: none (0–0.20), minimal (0.21–0.39), weak (0.40–0.59), moderate (0.60–0.79), strong (0.80–0.90), and almost perfect (greater than 0.90) (35). Bland–Altman plots were generated to analyze agreement in pain, agitation, and sedation scores among raters. Finally, Spearman correlation analysis was performed to examine relationships between pain, agitation, and sedation scores assigned by each rater, given that sedative use may suppress behavioral pain indicators without providing analgesia (22).

3 Results

3.1 Demographic characteristics of nurses

Nineteen NICU nurses participated in the study. Their characteristics are summarized in Table 1. The mean duration of NICU experience was 3.37 ± 2.19 years.

TABLE 1 Characteristics of NICU nurses (N = 19).

Variables of nurses	Mean (SD)	Min–Max (Median)
Age (years)	23.95 (1.65)	20–26 (24)
Duration of experience working in existing NICU (years)	3.37 (2.19)	1–6 (3)
Duration of total professional experience as NICU nurse (years)	4.58 (3.13)	1–7 (4)
Duration of experience using N-PASS scale	2.19 (0.61)	1–3 (2)
Sex	n	%
Female	18	94.7
Male	1	5.3
Educational level		
High school	6	31.6
Associate degree	2	10.5
Bachelor’s degree	11	57.9

3.2 Descriptive and clinical data of preterm infants

Descriptive and clinical characteristics of preterm infants are shown in Table 2. Most infants were male (57.4%). The mean gestational age was 30.86 ± 3.69 weeks, and the mean birth weight was 1,632.22 ± 636.29 g. Nearly half of the infants (48.9%) received respiratory support, and approximately one-fourth (24.2%) received sedation. In the previous 12 h, almost half of the infants (48.4%) received either pharmacologic or nonpharmacologic pain relief interventions.

3.3 Descriptive data for N-PASS

Descriptive statistics for the N-PASS are presented in Table 3. Among the 190 observations from 82 preterm infants, the mean pain and agitation scores was 3.00 ± 2.10 for the NICU nurses, 2.31 ± 1.38 for Researcher 1, and 2.29 ± 1.35 for Researcher 2. For the 46 infants who received sedation, the mean sedation scores were -1.52 ± 2.63 for the NICU nurses, -1.46 ± 2.87 for Researcher 1, and -1.37 ± 2.78 for Researcher 2. Regarding pain assessments, 65.3% of the nurses rated preterm infants’ pain as within an acceptable range. In contrast, both researchers rated 80% of the infants as having acceptable pain levels and 20% as having moderate pain. Among sedated infants, 71.9% of the nurses, 73.9% of rating from Researcher 1, and 84.8% of rating from Researcher 2 assigned a sedation score of 0.

3.4 Agreement level for N-PASS scores

The ICC among the three raters was evaluated using a two-way random-effects model with absolute agreement. Results are presented in Table 4. The single-measure ICC for N-PASS pain and agitation scores was 0.64 (95% CI, 0.57–0.71; F = 6.430; p < 0.001), and the average-measure ICC was 0.84 (95% CI, 0.80–0.88; F = 6.430; p < 0.001). For sedation scores among the 46 preterm infants receiving sedatives, the single-measure ICC was 0.84 (95% CI, 0.76–0.90; F = 16.812; p < 0.001) and the average-measure ICC was 0.94 (95% CI, 0.90–0.97; F = 16.812; p < 0.001).

TABLE 2 Demographic and clinical variables of preterm infants (N = 190).

Variables	Mean (SD)	Min–Max (Median)
Gestation week	30.86 (3.69)	24–36 (31.5)
Birth weight	1,623.22 (636.29)	530–3,010 (1,607.50)
Current weight	2033.07 (703.54)	530–4,900 (1,925)
Sex	n	%
Girl	81	42.6
Boy	109	57.4
Medical diagnosis		
Prematurity	162	85.3
Prematurity + respiratory distress + Apnea	18	9.5
Prematurity + bronchopneumonia	3	1.6
Prematurity + asphyxia	2	1.1
Prematurity + omphalitis	1	0.5
Prematurity + hyperbilirubinemia	1	0.5
Prematurity + sepsis + COVID-19	1	0.5
Prematurity + thrombocytopenia	1	0.5
Prematurity + fever	1	0.5
Receiving respiratory support		
Yes	93	48.9
No	97	51.1
Type of respiratory support (n = 93)		
HFO + intubation	34	36.5
Supplementation within the incubator	33	35.5
NCPAP	18	19.4
CPAP	8	8.6
Receiving sedation		
Yes	68	35.8
No	122	64.2
Interventions to reduce pain in the last 12 h		
Implemented	92	48.4
Not Implemented	98	51.6
Interventions for pain (n = 92)		
Nonpharmacologic methods		
Breast-feeding	21	22.8
Pacifier + position	18	19.6
Kangaroo	13	14.1
Pacifier + wrapping	11	12.0
Toot sweet	11	11.9
Positioning	9	9.8
Reduce touching	3	3.3
Touch + singing	2	2.2
Pharmacologic methods	4	4.3

Agreement levels for N-PASS pain subscale categories among the NICU nurses and the two researchers were analyzed using Fleiss kappa statistics (Table 5). The kappa values showed weak agreement for the no-pain category (κ = 0.49; 95% CI, 0.49–0.50; F = 11.783; p < 0.001) and the mild-pain category (κ = 0.44; 95% CI, 0.43–0.44; F = 10.434; p < 0.001). No agreement was observed for the severe-pain category (κ = 0.01; 95% CI, -0.02 to -0.01; F = 0.767), and this result was not statistically significant (p > 0.05). Because severe pain was rare in this sample, the kappa estimate for this category was statistically unstable and its CI should not be considered a reliable indicator of agreement.

TABLE 3 Descriptive and comparison statistics of neonatal pain, agitation, and sedation scale scores by NICU nurses and two researchers.

N-PASS	Raters	Mean	SD	Median	Min/max	Skewness	Kurtosis
Pain -Agitation (N = 190)	Nurse	3.00	2.10	3	0/11	1.07	1.07
	Researcher 1	2.31	1.38	2	0/6	0.60	-0.46
	Researcher 2	2.29	1.35	2	0/6	0.67	-0.38
Paired sample t-test	Nurse-Researcher 1: $t = 5.839, df = 189, p < 0.001$						
	Nurse-Researcher 2: $t = 5.767, df = 189, p < 0.001$						
	Researcher 1-Researcher 2: $t = 0.589, df = 189, p = 0.557, p > 0.05$						
Sedation (n = 46)	Nurse	-1.52	2.63	0	0/-8	-1.42	0.52
	Researcher 1	-1.46	2.87	0	0/-10	-2.02	3.27
	Researcher 2	-1.37	2.78	0	0/-10	-2.20	4.18
Wilcoxon signed-ranks test	Nurse-Researcher 1: $z = -0.401, p = 0.688, p > 0.05$						
	Nurse-Researcher 2: $z = -0.581, p = 0.562, p > 0.05$						
	Researcher 1-Researcher 2: $z = -1.342, p = 0.180, p > 0.05$						
Pain -Agitation (N = 190)	Severity of pain	0-3 point (no pain)		4-7 point (mild)		8-11 point (severe)	
		n	%	n	%	n	%
	Nurse	124	65.3	59	31.1	7	3.7
Researcher 1	152	80.0	38	20.0	-	-	
Researcher 2	152	80.0	38	20.0	-	-	
Sedation (n = 46)	Level of sedation	0 point (no sedation)		-1 to -5 point (light)		-6 to -10 point (deep)	
		n	%	n	%	n	%
	Nurse	33	71.7	8	17.4	5	10.9
Researcher 1	34	73.9	8	17.4	4	8.7	
Researcher 2	39	84.8	4	8.7	3	6.5	

TABLE 4 Intraclass correlation among raters for the total mean N-PASS pain, agitation, and sedation scores.

N-PASS pain, agitation scores (N = 190)	ICC	95% confidence interval		F test with true value 0			
		LB	UB	F	df1	df2	p
Single measure	0.64	0.57	0.71	6.430	189	378	<0.001
Average measure	0.84	0.80	0.88	6.430	189	378	<0.001
N-PASS Sedation Scores (n = 46)	ICC	95% confidence interval		F test with true value 0			
		LB	UB	F	df1	df2	p
Single measure	0.84	0.76	0.90	16.812	45	90	<0.001
Average measure	0.94	0.90	0.97	16.812	45	90	<0.001

ICC, intraclass correlation; df, degree of freedom; LB, lower bound; UB, upper bound.

For sedation categories, moderate agreement was observed for the no-sedation group ($\kappa = 0.76$; 95% CI, 0.75-0.76; $F = 8.880$; $p < 0.001$). Weak agreement was found for the light-sedation group ($\kappa = 0.53$; 95% CI, 0.53-0.54; $F = 6.252$; $p < 0.001$) and the deep-sedation group ($\kappa = 0.45$; 95% CI, 0.45-0.46; $F = 5.314$; $p < 0.001$).

3.5 Bland-Altman graphics

Bland-Altman plots for the pain/agitation and sedation subscales are shown Figures 1, 2. The average differences between the nurse and both researchers were small. Most score differences fell within the 95% CI, indicating measurement

consistency within clinically acceptable limits for pain/agitation assessments. The Bland-Altman analysis for sedation scores also demonstrated extremely strong agreement among the raters.

3.6 Correlations between each rater's pain/agitation and sedation scores

Spearman correlation analyses were conducted to examine the relationships between each rater's pain/agitation and sedation scores for individual preterm infants, and the results are presented in Supplementary Table S1. Across all 190 observations from 82 preterm infants, the correlation coefficients

TABLE 5 Inter-rater reliability of N-PASS pain and sedation scores using Fleiss Kappa.

N-PASS pain, agitation scores (<i>N</i> = 190) ^a	Conditional probability	Kappa	SE	<i>z</i>	<i>p</i>	95% confidence interval	
						LB	UB
Severity of pain							
Agreement for “0–3 point (no pain)”	0.75	0.49	0.04	11.783	<0.001	0.49	0.50
Agreement for “4–7 point (mild)”	0.24	0.44	0.04	10.434	<0.001	0.43	0.44
Agreement for “8–11 point (severe)”	0.01	−0.01	0.04	−0.297	0.767	−0.02	−0.01
Overall agreement		0.45	0.04	11.280	<0.001	0.44	0.45
N-PASS sedation scores (<i>n</i> = 46) ^b	Conditional probability	Kappa	SE	<i>z</i>	<i>p</i>	95% confidence interval	
						LB	UB
Level of sedation							
Agreement for “0 point (no sedation)”	0.77	0.76	0.09	8.880	<0.001	0.75	0.76
Agreement for “1 to −5 point (light)”	0.15	0.53	0.09	6.252	<0.001	0.53	0.54
Agreement for “−6 to −10 point (deep)”	0.09	0.45	0.09	5.314	<0.001	0.45	0.46
Overall agreement		0.62	0.07	9.397	<0.001	0.61	0.62

LB, lower bound; UB, upper bound.

^aSample data contains 190 effective subjects and 3 raters.

^bSample data contains 46 effective subjects and 3 raters.

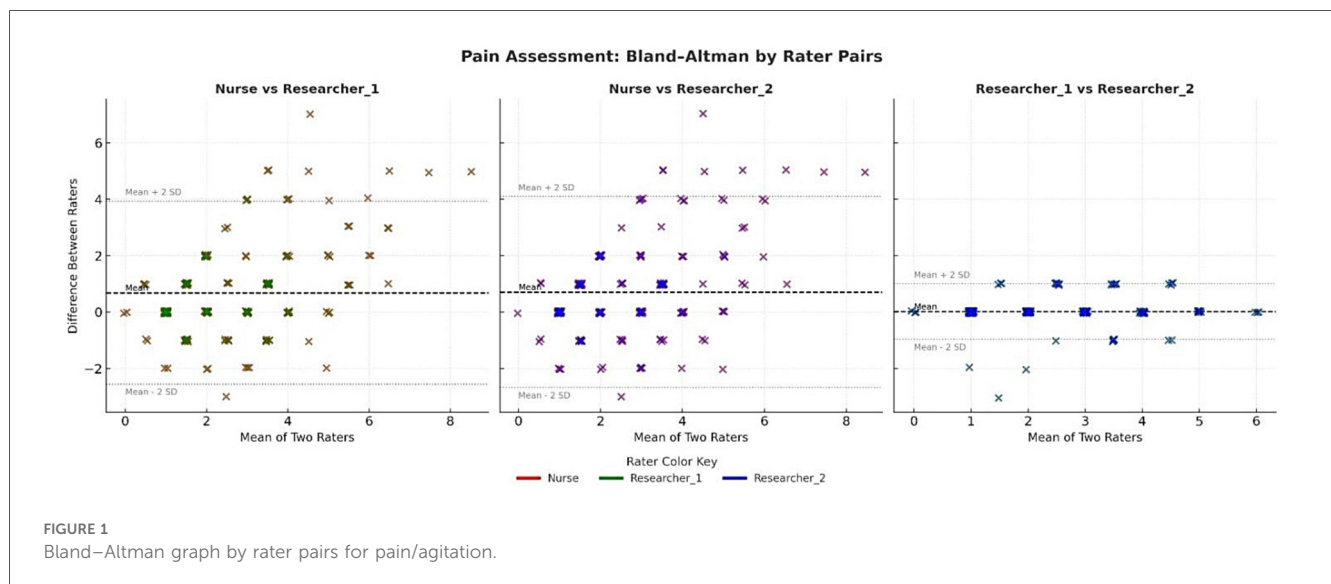


FIGURE 1 Bland–Altman graph by rater pairs for pain/agitation.

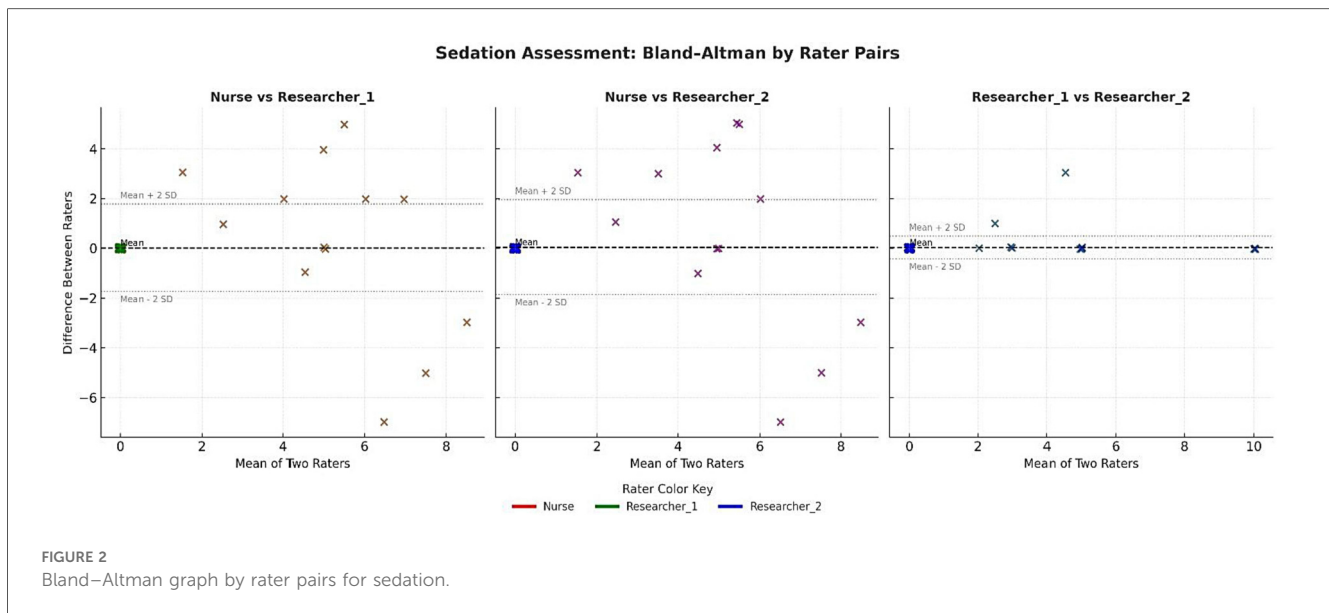
between pain/agitation and sedation scores were $r = -0.03$ ($p > 0.05$) for nurses, $r = -0.14$ ($p < 0.05$) for Researcher 1, and $r = -0.13$ ($p > 0.05$) for Researcher 2. Among infants who received sedation, the correlation coefficients were $r = -0.01$ ($p > 0.05$) for nurses, $r = -0.31$ ($p < 0.05$) for Researcher 1, and $r = -0.24$ ($p > 0.05$) for Researcher 2. For infants who received analgesic medications or nonpharmacologic interventions, the correlation coefficients were $r = -0.01$ ($p > 0.05$) for nurses, $r = -0.22$ ($p < 0.05$) for Researcher 1, and $r = -0.18$ ($p > 0.05$) for Researcher 2.

4 Discussion

This study evaluated the inter-rater reliability of the N-PASS in a real-time clinical environment. The findings provide insight

into how consistently pain/agitation and sedation scores are applied by different assessors during routine NICU care.

The findings of this study revealed that inter-rater reliability for the pain/agitation subscale was moderate based on a single measurement (ICC = 0.64) but improved to “good” when averaged across raters (ICC = 0.84). These results indicate that pain assessment is subject to observer interpretation; however, scoring consistency may improve with the number of raters. Our results are lower than those reported in existing literature, which consistently indicates higher reliability. Previous research has consistently exhibited higher inter-rater reliability for the N-PASS than our real-time clinical assessments. Hummel et al. (2008) reported ICC values >0.90 under structured conditions with trained nurses, and Hummel (2017) similarly found reliability values between 0.86 and 0.93 in preterm and older infants (22, 25). Video-based assessments by Huang and



Kappesser also demonstrated excellent reliability (ICC >0.90) during both painful and nonpainful procedures (32, 34). More recently, Benbrook et al. (2023) reported good reliability (ICC = 0.83–0.85) among bedside nurses following structured N-PASS training (23). The higher reliability observed in these studies likely reflects methodological differences, such as the use of trained raters, controlled environments, and standardized assessment protocols. These factors differ substantially from the real-time bedside conditions of the present study.

In this study, the kappa analysis results for inter-rater reliability on the pain/agitation subscale revealed poor agreement in the “no-pain” ($\kappa = 0.49$) and “mild-pain” ($\kappa = 0.44$) categories and no agreement in the “severe-pain” category ($\kappa = -0.01$). Therefore, raters expressed differing opinions when identifying severe pain in infants. As seen in Table 3 and Figure 1, nurses consistently assigned higher pain scores than the researchers, particularly in the mild-to-moderate range, which contributed to poor kappa agreement in these categories. Cignacco et al. (2008) reported that nurses perceived procedures performed on infants as more painful than physicians did (36). Such overestimation may increase the likelihood of unnecessary pharmacologic interventions, whereas underestimation may lead to insufficient analgesia, both of which have important clinical implications and may hinder opioid-sparing strategies in the NICU (37, 38). This variability in pain scoring also carries therapeutic implications (39).

Interpreting kappa coefficients in this study requires caution due to the highly unbalanced distribution of pain categories. The near absence of infants rated with severe pain created a sparse-data problem known to lower kappa scores and produce mathematically unstable estimates. This imbalance also explains the artificially narrow CIs, which result from low variance rather than high measurement precision. Therefore, lower kappa values in this context do not necessarily indicate true disagreement among raters but rather reflect a statistical limitation caused by

categorical distribution. Under these conditions, ICC values derived from continuous N-PASS scores offer a more accurate representation of inter-rater reliability.

Considering that pain assessment is a core parameter evaluated at least every 4 h in the NICU, consistent scoring by bedside nurses is essential to guide appropriate pharmacologic and nonpharmacologic interventions (40). Therefore, strengthening training and standardization efforts may enhance inter-rater reliability and support safer, more effective pain management practices in preterm infants.

The findings of this study indicated good reliability (ICC = 0.84) for single measures and excellent reliability (ICC = 0.94) for average measures on the sedation subscale, which are consistent with previous research results (23, 25). Furthermore, in the study conducted by (23), nurses reported that the N-PASS effectively assessed neonatal sedation status at a “good” or “very good” level. The findings of this study revealed good reliability for single measures (ICC = 0.84) and excellent reliability for average measures (ICC = 0.94) on the sedation subscale, consistent with previous research (23, 25). In the present analysis, the mean kappa value among three raters indicated moderate agreement, and Figure 2 demonstrates a close distribution of sedation scores between the nurses and researchers, in contrast to the wider dispersion observed in pain scoring. This pattern may reflect that sedation cues tend to be interpreted more consistently than pain-related behaviors.

Another notable finding in this study is that researchers identified a negative correlation between pain/agitation and sedation scores for the same patient, a finding consistent with known pharmacologic effects where sedatives may reduce observable pain behaviors and analgesics may contribute to mild sedation (22). In contrast, this relationship was not evident in nurse-assigned scores, particularly among infants receiving sedatives or analgesics. Furthermore, nurses showed less consistency when rating values above zero, often assigning

higher sedation scores. These discrepancies suggest that distinguishing overlapping pain- and sedation-related cues is more challenging during routine bedside care, whereas investigator assessments exhibited a more consistent scoring pattern.

Although nurse characteristics could not be examined separately in this study, previous research indicates that demographic factors and clinical experience do not account for variability in pain scoring and that standardized education and structured approaches are more influential in improving inter-rater consistency (36).

Limitations of this study include: 1) First, the study's single-center conduct may be a limitation. However, the participation of approximately three-quarters of the nurses working on the unit, the fact that raters assessed pain and sedation in real clinical setting, and the selection of an appropriate method for ICC analysis may support the study's generalizability. Second, because data collection occurred in a real clinical setting and researchers faced time constraints, all data were gathered only during the day shift, which may have introduced selection bias. A multicenter design and data collection across both day and night shifts are recommended for future studies. Third, although daytime-only data collection is not expected to substantially affect sample representativeness, a minor possibility of selection bias cannot be entirely ruled out.

5 Conclusion

This prospective, blinded observational study examined the inter-rater agreement of the N-PASS. In total, 190 observations from 82 preterm infants were evaluated by two researchers and 19 NICU nurses. The findings demonstrated varying levels of agreement across N-PASS subscales. The ICC values indicated moderate-to-good reliability for the pain/agitation subscale and excellent reliability for the sedation subscale. Consistent with the ICC results, the Kappa analysis showed lower agreement for pain/agitation but higher agreement for sedation. Notably, nurses assigned significantly higher pain scores than the researchers.

To reduce this variability and strengthen clinical care, a multidisciplinary approach emphasizing enhanced education is recommended. NICU nurses should focus on both theoretical knowledge and practical skill development through innovative training methods, such as simulation or video-based calibration. Furthermore, collaborative refinement of pain and sedation assessment and management protocols is essential to promote consistency and alignment across all healthcare disciplines in the NICU.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was conducted according to the guidelines of the Declaration of Helsinki, and all procedures involving human subjects were approved by the "Acibadem Mehmet Aydinlar University Medical Research Review Board (ATADEK) (Decision no: 2022-20/25, December 30, 2022). Written and verbal consent were obtained from the nurses participating in the study.

Author contributions

SA: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. ZKÖ: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft. EŞ: Conceptualization, Data curation, Writing – original draft. MY: Conceptualization, Data curation, Writing – original draft.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

We would like to thank all nurses who participated in this study. We also gratefully acknowledge Atilla Bozdoğan for his support and assistance in analyzing the study's statistical data.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Campbell-Yeo M, Eriksson M, Benoit B. Assessment and management of pain in preterm infants: a practice update. *Children* (2022) 9(244):1–18. doi: 10.3390/children9020244
- Orovec A, Disher T, Caddell K, Campbell-Yeo M. Assessment and management of procedural pain during the entire neonatal intensive care unit hospitalization. *Pain Manag Nurs* (2019) 20(5):503–11. doi: 10.1016/j.pmn.2018.11.061
- Luo F, Zhu H, Mei L, Shu Q, Cheng X, Chen X, et al. Evaluation of procedural pain for neonates in a neonatal intensive care unit: a single-centre study. *BMJ Paediatrics Open* (2023) 7(1): e002107. doi: 10.1136/bmjpo-2023-002107
- Williams MD, Lascelles BD. Early neonatal pain review of clinical and experimental implications on painful conditions later in life. *Front Pediatr* (2020) 8:30. doi: 10.3389/fped.2020.00030
- Boggini T, Pozzoli S, Schiavolin P, Erario R, Mosca F, Brambilla P, et al. Cumulative procedural pain and brain development in very preterm infants: a systematic review of clinical and preclinical studies. *Neurosci Biobehav Rev* (2021) 123: 320–36. doi: 10.1016/j.neubiorev.2020.12.016
- Doesburg SM, Chau CM, Cheung TP, Moiseev A, Ribary U, Herdman AT, et al. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *J Int Assoc Study Pain* (2013) 154(10):1946–52. doi: 10.1016/j.pain.2013.04.009
- Pavlyshyn H, Sarapuk I, Kozak K. The relationship between neonatal stress in preterm infants and developmental outcomes at the corrected age of 24–30 months. *Front Psychol* (2024) 15:1415054. doi: 10.3389/fpsyg.2024.1415054
- Ranger M, Chau CM, Garg A, Woodward TS, Beg MF, Bjornson B, et al. Neonatal pain-related stress predicts cortical thickness at age 7 years in children born very preterm. *PLoS One* (2013) 8(10): e76702. doi: 10.1371/journal.pone.0076702
- Pereira FL, Gasparido CM. Neonatal pain and developmental outcomes in children born preterm: an updated systematic review. *Psychol Neurosci* (2024) 17(1): 1–15. doi: 10.1037/pne0000332
- Perry M, Tan Z, Chen J, Weidig T, Xu W, Cong XS. Neonatal pain: perceptions and current practice. *Crit Care Nurs Clin* (2018) 30(4):549–61. doi: 10.1016/j.cnc.2018.07.013
- Victoria NC, Murphy AZ. Exposure to early life pain: long term consequences and contributing mechanisms. *Curr Opin Behav Sci* (2016) 7:61–8. doi: 10.1016/j.cobeha.2015.11.015
- Carlsen Mistic M, Andersen RD, Strand S, Eriksson M, Olsson E. Nurses' perception, knowledge, and use of neonatal pain assessment. *Pediatr Neonat Pain* (2021) 3(2):59–65. doi: 10.1002/pne2.12050
- Yiğit Ş, Ecevit A, Köröglü ÖA. Turkish neonatal society guideline on the neonatal pain and its management. *Turk Arch Pediatr* (2018) 53(1):161–71. doi: 10.5152/TurkPediatriArs.2018.01802
- Llerena A, Tran K, Choudhary D, Hausmann J, Goldgof D, Sun Y, et al. Neonatal pain assessment: do we have the right tools? *Front Pediatr* (2023) 10: 1022751. doi: 10.3389/fped.2022.1022751
- Leena-Mari H, Riikka L. Pain in premature infants?: Pain reactions and possible consequences in later life (Thesis). (2016). Available online at: <https://www.theseus.fi/handle/10024/120228> (Accessed March 15, 2025).
- Pölkki T, Korhonen A, Axelin A, Saarela T, Laukkala H. Development and preliminary validation of the neonatal infant acute pain assessment scale (NIAPAS). *Int J Nurs Stud* (2014) 51(12):1585–94. doi: 10.1016/j.ijnurstu.2014.04.001
- Fitri SYR, Lusmilasari L, Juffrie M, Rakhmawati W. Pain in neonates: a concept analysis. *Anesth Pain Med* (2019) 9(4), e92455. doi: 10.5812/aapm.92455
- Arabiat D, Mörelius E, Hoti K, Hughes J. Pain assessment tools for use in infants: a meta-review. *BMC Pediatr* (2023) 23(1):1–22. doi: 10.1186/s12887-023-04099-7
- Anand KJ, Eriksson, Boyle EM, Avila-Alvarez A, Andersen RD, Sarafidis K, et al. EUROPAIN survey working group of the NeoOpioid consortium. Assessment of continuous pain in newborns admitted to NICUs in 18 European countries. *Acta Paediatr* (2017) 106(8):1248–59. doi: 10.1111/apa.13810
- García-Rodríguez MT, Bujan-Bravo S, Seijo-Bestilleiro R, Gonzalez-Martin C. Pain assessment and management in the newborn: a systematized review. *World J Clin Cases* (2021) 9(21):5921–31. doi: 10.12998/wjcc.v9.i21.5921
- Olsson E, Ahl H, Bengtsson K, Vejayaram DN, Norman E, Bruschetti M, et al. The use and reporting of neonatal pain scales: a systematic review of randomized trials. *Pain* (2021) 162(2):353–60. doi: 10.1097/j.pain.0000000000002046
- Hummel P. Psychometric evaluation of the neonatal pain, agitation, and sedation (N-PASS) scale in infants and children up to age 36 months. *Pediatr Nurs* (2017) 43(4):175–84.
- Benbrook K, Manworren RC, Zuravel R, Entler A, Riendeau K, Myler C, et al. Agreement of the neonatal pain, agitation, and sedation scale (N-PASS) with NICU nurses' assessments. *Adv Neonatal Care* (2023) 23(2):173–81. doi: 10.1097/ANC.0000000000000968
- Hillman BA, Tabrizi MN, Gauda EB, Carson KA, Aucutt SW. The neonatal pain, agitation and sedation scale and the bedside nurse's assessment of neonates. *J Perinatol* (2015) 35(2):128–31. doi: 10.1038/jp.2014.154
- Hummel P, Puchalski M, Creech SD, Weiss MG. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol* (2008) 28(1):55–60. doi: 10.1038/sj.jp.7211861
- Von Elm E, Altman DG, Egger M, Pocock SJ, Göttsche PC, Vandenbroucke JP. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* (2007) 370(9596):1453–7. doi: 10.1016/S0140-6736(07)61602-X
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* (2016) 15(2): 155–63. doi: 10.1016/j.jcm.2016.02.012
- Giordano V, Deindl P, Kuttner S, Waldhör T, Berger A, Olischar M. The neonatal pain, agitation and sedation scale reliably detected oversedation but failed to differentiate between other sedation levels. *Acta Paediatr* (2014) 103(12): e515–21. doi: 10.1111/apa.12770
- Morgan ME, Kukora S, Nemshak M, Shuman CJ. Neonatal pain, agitation, and sedation scale's use, reliability, and validity: a systematic review. *J Perinatol* (2020) 40:1699–710. doi: 10.1038/s41372-020-00840-7
- Lin Z, Zheng X, He R, Li X, Shen Q, Meng Y, et al. Reliability and validity of LAN's neonatal pain scale for acute pain assessment in neonates with mechanical ventilation. *Chin Nurs Res* (2020) 34(21):3801–6. doi: 10.12102/j.issn.1009-6493.2020.21.011
- Açıkgöz A, Çiğdem Z, Yıldız S, Demirüstü C, Yazar M, Akşit A. Turkish adaptation of the neonatal pain/agitation, sedation scale (N-PASS) and its validity and reliability. *Indian J Fundam Appl Life Sci* (2017) 7(2):5–11.
- Huang XZ, Li L, Zhou J, He F, Zhong CX, Wang B. Evaluation of three pain assessment scales used for ventilated neonates. *J Clin Nurs* (2018) 27(19–20):3522–9. doi: 10.1111/jocn.14585
- Hummel P, Lawlor-Klean P, Weiss MG. Validity and reliability of the N-PASS assessment tool with acute pain. *J Perinatol* (2010) 30(7):474–8. doi: 10.1038/jp.2009.185
- Kappesser J, Kamper-Fuhrmann E, de Laffolie J, Faas D, Ehrhardt H, Franck LS, et al. Pain-specific reactions or indicators of a general stress response?: investigating the discriminant validity of 5 well-established neonatal pain assessment tools. *Clin J Pain* (2019) 35(2):101–10. doi: 10.1097/AJP.0000000000000660
- McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* (2012) 22(3):276–82. doi: 10.1016/j.jcm.2016.02.012
- Cignacco E, Hamers JP, Stoffel L, Zimmermann LJI. Routine procedures in NICUs: factors influencing pain assessment and ranking by pain intensity. *Swiss Med Wkly* (2008) 138(3334):484. doi: 10.4414/sm.w.2008.12147
- McPherson C, Ortinau CM, Vesoulis Z. Practical approaches to sedation and analgesia in the newborn. *J Perinatol* (2021) 41(3):383–95. doi: 10.1038/s41372-020-00878-7
- McPherson C, Grunau RE. Pharmacologic analgesia and sedation in neonates. *Clin Perinatol* (2022) 49(1):243–65. doi: 10.1016/j.clp.2021.11.014
- Cocchi E, Shabani J, Aceti A, Ancora G, Corvaglia L, Marchetti F. Dexmedetomidine as a promising neuroprotective sedoanalgesic in neonatal therapeutic hypothermia: a systematic review and meta-analysis. *Neonatology* (2025) 122(4):495–504. doi: 10.1159/000546017
- Mala O, Forster EM, Kain VJ, Thai nurses' and midwives' perceptions regarding barriers, facilitators, and competence in neonatal pain management. *Adv Neonatal Care* (2024) 24(2): E26–38. doi: 10.1097/ANC.0000000000001128