Journal of Pediatric Nursing 48 (2019) e49–e55

Contents lists available at ScienceDirect

Journal of Pediatric Nursing

journal homepage: www.pediatricnursing.org

Psychometric Testing of the Turkish Version of the Premature Infant Pain Profile Revised-PIPP-R

Ayşe Şener Taplak [a,*], Meral Bayat [b]

[a] Yozgat Bozok University, Faculty of Health Sciences, Department of Pediatric Nursing, Turkey
[b] Kayseri Erciyes University, Faculty of Health Sciences, Department of Pediatric Nursing, Turkey

ARTICLE INFO

Article history:
Received 19 January 2019
Revised 8 June 2019
Accepted 8 June 2019

Keywords:
PIPP-R
Validity and reliability
Selectivity and sensitivity

ABSTRACT

This methodological study was conducted to determine the Turkish validity-reliability and selectivity-sensitivity of the Premature Infant Pain Profile-Revised Form (PIPP-R). 200 newborn infants with gestational age of 26–42 weeks followed up in the neonatal unit of a university hospital. Necessary permissions were obtained from the responsible author, the institution, and the ethics committee before starting the study. Language validity, content validity, and construct validity studies were conducted for the validity of PIPP-R. For the content validity, the opinions of 10 experts were asked and the content validity index was found as 0.88. The factor analysis method was used to determine the construct validity of the scale and it was determined that the PIPP-R consisted of three factors. The Cronbach's alpha coefficient from the internal consistency analysis was calculated for the reliability of the scale and the scale was determined to be highly reliable with the value of 0.840. The intra-class correlation coefficient was used in determining the observer reliability and the agreement between three observers was found to be very good (0.944–1.000). In the analyses conducted for the sensitivity and selectivity of the scale, it was determined that the sensitivity of the scale was 91% and its selectivity was 88%. In accordance with the obtained results, PIPP-R was determined to be a valid, reliable, selective, and sensitive measurement tool to be used for the pain assessment of the Turkish population.

Introduction

Recent pain studies have indicated that an infant admitted to the neonatal intensive care unit is routinely exposed to, on average, 10–16 painful procedures daily (Carbajal, Rousset, Danan, et al., 2008; Roofthooft, Simons, Anand, Tibboel, & van Dijk, 2014; Stevens, McGrath, Gibbins, et al., 2003). In addition to the stressful clinical setting, the recurrent pain experienced for the purpose of diagnosis and treatment leads to atrophy in the brain of the infant, which can result in neurodevelopmental problems due to intraventricular hemorrhage and a decrease in the subcortical white and gray matter in the brain (Aarnoudse-Moens, Weisglas-Kuperus, van Goudoever, & Oosterlaan, 2009; Brummelte, Grunau, Chau, et al., 2012; Hall & Anand, 2005). The prevention or minimization of pain is the right of any newborn infant. As infants experience a healthy neonatal period when their pain is under control, their duration of stay in hospital is reduced and their growth and development can accelerate, which contributes positively to the national economy (Akcan & Polat, 2017; Eroğlu & Arslan, 2018).

Effective pain management is required to eliminate the adverse effects of pain. The most important step in the management of pain associated with ınterventional procedures is the assessment of pain. However, the greatest challenge encountered in pain assessment in infants is that pain cannot be expressed verbally (Akcan & Polat, 2017; Carter & Brunkhorst, 2017; Eroğlu & Arslan, 2018). Despite this, it is possible to interpret the indicators of pain in infants and minimize the negative effect of pain. For this purpose, various scales, including behavioral and physiological parameters that can be used to identify the pain intensity in infants, have been developed (Eroğlu & Arslan, 2018). There are few scales adapted to the Turkish population among the scales used in the assessment of pain in infants. One such scale is the Premature Infant Pain Profile (PIPP) (Stevens, Johnston, Petryshen, & Taddio, 1996). This scale has been used in many studies in Turkey (Akcan & Yiğit, 2015; Dolgun & Bozlak, 2017; Kabataş et al., 2016; Küçük Alemdar & Kardaş Özdemir, 2017; Şener Taplak & Erdem, 2017); however, problems have been encountered in the clinical use of the PIPP, including the fact it does not cover infants who are born at <28 weeks or >36 weeks, and that the pain score can be exaggerated as a result of the scoring system. Therefore, this scale was revised by Stevens, Gibbins, Yamada, et al. (2014) as the PIPP-Revised form (PIPP-R), which included revisions of the scoring system and extending the applicable gestational age range (Stevens et al., 2014).

* Corresponding author.
E-mail addresses: hem.ayse_3886@hotmail.com, ayse.taplak@bozok.edu.tr (A.Ş. Taplak).

https://doi.org/10.1016/j.pedn.2019.06.007
0882-5963/

The PIPP-R, which is an updated version of the PIPP, is a multidimensional pain assessment tool. Since the item statements in the scale were revised to improve their comprehensibility, pain assessment in disadvantaged groups can be more objective owing to the revised scoring system, the wide range of gestational ages suitable for its use in pain assessment (covering the period <28 weeks and >37 weeks), and the application being easier. It was determined that, compared with the PIPP, the clinical benefits of PIPP-R were increased as a result of these beneficial changes (Gibbins, Stevens, Yamada, et al., 2014; Stevens et al., 2014).

Aim

This study was conducted to determine the validity, reliability, selectivity and sensitivity of the PIPP-R in the Turkish population, which has become a more comprehensive and effective assessment tool for pain assessment in infants.

Methods

Study design

This methodological study was performed to determine the validity, reliability, selectivity and sensitivity of the PIPP-R in the Turkish population.

Population and sample

The population examined in the present study included infants who were followed up in the neonatal unit of a university hospital between August 2015 and June 2016. Those infants with a gestational age of ≥26 weeks, a postnatal age of 3–5 days, who underwent heel lance for phenylketonuri screening (PKU), were able to tolerate enteral nutrition (absence of necrotizing enterocolitis, digestive system or chromosome abnormalities), did not receive any analgesic/sedative drug treatment, or any medical treatment other than the appropriate vitamin supplements and antibiotic treatments, were included in the study. Those infants connected to a mechanical ventilator, with congenital anomalies or that had undergone surgical intervention were excluded from the study.

To determine the validity and reliability of the scale, the sample size should be at least 10–20 times greater than the item number on the scale (Alpar, 2016; Çapık, 2014). In this respect, the painful procedure (heel lance for PKU screening) was performed in 250 newborn infants for PIPP-R with seven items, who met the inclusion criteria, which was recorded using a video camera. Additionally, a painless procedure (diaper change) in these infants was recorded using the video camera in the same shift to enable selectivity and sensitivity analysis. These records were then evaluated by three observers in a double-blind manner. A total of 50 newborn infants was considered by the observers as ineligible for the evaluations and were excluded from the sample group for the following reasons: the pulse oximetry device was not fully visible, low quality video records, video started directly from the heel lance without a 15-Sec observation period during evaluation. The diaper changing records of these infants was also excluded from the study.

As a result, the sample group in the study included a total of 200 newborn infants, who were subjected to the heel lance and diaper changing procedures. In the analysis performed to determine the adequacy of the sample size for the validity and reliability analyses, it was found that the sample size was adequate for the Kaiser-Meyer-Olkin (KMO 0.80) and Bartlett's test (0.000).

Data collection tools

The Infant Evaluation Data Form and the PIPP-R were used for data collection. The Infant Evaluation Data Form compressed data, including the descriptive characteristics (gender, gestational age, natal age, and type of birth) and physical measurements of ınfants (birth and body weight). This information was obtained from the infant's mother and from patient files.

The PIPP-R is a Likert type scale. In the scoring of the scale, items related to the physiological and behavioral indicators score as 0, 1, 2, and 3 points for each variable, reflecting the difference between the values obtained at the beginning and during the process. The items related to the contextual indicators (behavioral state and gestational age) score as 3, 2, 1 and 0 points at the beginning of the pain assessment (before contact with the infant). According to the PIPP-R, the pain of the infant is evaluated over the total score. Behavioral and physiological indicators are scored at the end of the scale in scoring of the PIPP-R. Accordingly, if the infant scores zero points (subtotal) from the overall physiological and behavioral indicators, the contextual indicators (behavioral state and gestational age) are not scored. If the infant receives a score above zero from the baseline parameters, the total score is obtained by including the contextual indicators in the scoring. This prevents an erroneous increase of the pain score based on the baseline static variables when no pain response is observed in the infant. Accordingly, the highest score of the PIPP-R is 21 for preterm infants and 18 for term infants (Stevens et al., 2014). Stevens et al. stated that the pain experienced by infants is mild if the mean PIPP-R is 0–6 points, moderate if the mean is 7–12 points, and severe if the mean is 13–21 points (Stevens et al., 2014).

Study protocol

The study was performed in two stages, comprising a cultural adaptation process and psychometric assessment of the PIPP-R.

Cultural adaptation process

The cultural adaptation process of the scale was performed in two stages: language validity and content validity. In the language validity stage of the PIPP-R, the scale was first translated from English, being the original language, into Turkish. The scale items were translated independently into Turkish by three experts who were experienced in the neonatal field and had a good command of the two different languages. The Turkish version of the scale was successfully prepared, following which the scale was translated back from Turkish into English by three experts who had a good command of both languages. The English translations of the scale were performed by an expert who spoke good English. The translated version of the scale was compared with its original version. The language validity of the scale was completed in accordance with the expert opinions.

The Lawshe's technique was used for the content validity of the scale, the opinions of a total of 10 experts in the pediatrics department and pediatric nursing department were obtained. These individuals were asked to score the items of the scale as appropriate, the item should be reviewed slightly, the item should be considerably reviewed, and inappropriate. As a result of the evaluations, the Content Validity Rate (CVR), and Content Validity Index (CVI) were evaluated for each item on the scale (Alpar, 2016; Yurdugül, 2005). The CVI was obtained from the mean values of the CVR. In the evaluations made by the 10 experts, a CVR of >0.60 indicated a good level of validity (Yurdugül, 2005).

Psychometric assessment of the PIPP-R

The validity and reliability study of the PIPP-R was conducted by recording the heel lance procedure applied to 200 newborn infants for the PKU screening using a video camera. The diaper changing process of the infants subjected to a heel lance was also recorded using the video camera during the same shift, and these records were used for determining the selectivity and sensitivity of the scale.

In the neonatal unit, the heel lance procedure for PKU screening of the infants was performed by a nurse responsible for only this

procedure. For this reason, the heel lance procedure was performed by the same individual. The diaper changing procedure was performed by a nurse who was working during the entire day shift. The heel lance and diaper changing procedures were performed between 9 and 12 am in the same shift. Upon completion of the video recording of the infants, the videos were evaluated by three observers, who were specialized in the field, independent of each other. The evaluations were double-blind and the observers were not informed of the criteria of the procedure (painful or painless).

Validity

The most common method used for determining the construct validity of a scale is a factor analysis. For this purpose, whether or not the items in the test were suitable for factor analysis, and whether or not the sample size was adequate for factor analysis were evaluated. Accordingly, the adjusted total item correlation values, KMO and Bartlett's tests were used, as they are among the main tests evaluating the suitability of the PIPP-R for factor analysis. It was determined that the KMO was 0.803 > 0.50 and Bartlett's value was $\chi^2 = 719$; $p < 0.001$ for the PIPP-R and the sample size was considered adequate (KMO = 0.80–0.89 indicates a sample size is good). The adjusted total item correlation values from the main assessments for evaluating the suitability of the PIPP-R for factor analysis were calculated, and the adjusted total item correlation values for each scale item were >0.40. The PIPP-R was considered to be adequate for factor analysis. Principal component analysis and varimax rotation methods were used to examine the factor structure (Büyüköztürk, 2015). Generally, in factor analysis, the factors with Eigenvalues of ≥1 are considered important factors. The view, stating that the number of Eigenvalues explaining 65% of the total variance should be taken into consideration was used (Alpar, 2016).

Reliability assessment

The alpha coefficient is often used to calculate the reliability of a Likert type scale prepared over total scores. A high alpha coefficient of a scale indicates that the items in the scale work together at the same ratio. The scale is evaluated as highly reliable at alpha coefficients of 0.80–1.00, very reliable as those of 0.60–0.79, low reliability as those of 0.40–0.59, and unreliable at 0.00–0.39 (Alpar, 2016). This reliability criterion is used in cases in which multiple observers measure the same situation independently from each other using the same measurement tool. In this study, the intraclass correlation coefficient (ICC), special variance analysis, and Cronbach's alpha coefficient were used in order to evaluate the agreement between the independent observers. A consistency ≥70% between the observers was considered as suitable in the reliability assessment (Karakoç & Dönmez, 2014).

Table 1
Descriptive characteristics of infants.

Descriptive characteristics	Number	Percentage
Gestational age (week)		
26–31 weeks	60	30.0
32–36 weeks	71	35.5
37–42 weeks	69	34.5
Natal age (day)		
3 day	92	46.0
4 day	75	37.5
5 day	33	16.5
Gender		
Female	88	44.0
Male	112	56.0
Birth weight (gr)		
<1500 g	44	22.0
≥1501 g	156	78.0
Body weight (gr)		
<1500 g	38	19.0
≥1501 g	162	81.0
Type of birth		
Cesarean	149	74.5
Normal birth	51	25.5

Table 2
CVR and CVI values of the items of PIPP-R.

Scale items	Content Validity Rate (CVR)
Change in heart rate	0.8
Decrease in oxygen saturation	0.8
Brow bulge	0.8
Eye squeeze	1.0
Naso-labial furrow	0.8
Gestastional age	1.0
Baseline behavioral state	1.0
Content Validity Index (CVI)	**0.88**

Selectivity and sensitivity

Where selectivity determines how much pain is experienced in a painful procedure, the sensitivity allows for the determination of how much pain is experienced in a painless procedure. For this purpose, the sensitivity and selectivity formulas were utilized (Alpar, 2016). In addition, Stevens et al. stated that testing the extreme groups of a scale is one of the most effective methods in determining its validity (Stevens et al., 1996; Stevens et al., 2014). Previous studies have used this method for determining both the validity and the selectivity and sensitivity of PIPP scale (Jonsdottir & Kristjansdottir, 2005; Stevens et al., 1996; Vederhus, Eide, & Natvig, 2006). As well as PIPP, PIPP-R was evaluated in painful and painless procedures and in different gestational age groups (Gibbins et al., 2014; Sadeghi et al., 2017; Stevens et al., 2014). In this study, the scale was also applied to the extreme groups (different gestational age groups; painful and painless procedure) to determine the selectivity and sensitivity of the PIPP-R. The mean PIPP-R scores of the infants were assigned into three groups according to their gestational ages (26–31, 32–36, and 37–42 weeks); their mean painful and painless procedure scores were compared to determine the selectivity and sensitivity.

Statistical analysis

The IBM SPSS Statistics 21.0 (Chicago, IL) package was used for the data analysis. In data assessment, the number, percentage, mean, and standard deviation were used in the statistical analysis of descriptive data. The intraclass correlation coefficient was used to evaluate the agreement between the three observers, and the adjusted total item correlations obtained using a two-way mixed effect model were used to weighted the observers. The KMO and Bartlett's test of sphericity test were used to determine the adequacy of the sample size. Principal component analysis, Kaiser normalization, and Varimax rotation methods in factor analysis; selectivity, and sensitivity formulas, and One-way analysis of variance and Paired *t*-tests were used to determine the selectivity and sensitivity of the scale. $P < 0.05$ was considered to indicate a statistically significant difference.

Table 3
Explained total variance values.

Component	Eigenvalue	Variance explained by each component	Cumulative variance
1	2.548	36.40	36.40
2	1.651	23.58	59.98
3	1.602	22.89	**82.87**

Table 4
Matrix of converted components.

Scale Items	Factor 1	Factor 2	Factor 3
Brow Bulge	**0.887**	0.245	0.079
Eye Squeeze	**0.878**	0.219	0.210
Nasolabial Furrow	**0.858**	0.147	0.321
Change in Heart Rate	0.118	**0.868**	0.208
Decrease in Oxygen Saturation	0.390	**0.775**	0.024
Gestastional Age	0.266	-0.034	**0.883**
Baseline Behavioural State	0.135	0.407	**0.791**

The video recordings of the 200 infants (heel lance and diaper changing procedures) were evaluated separately according to the PIPP-R by three independent expert observers, including one pediatric nurse (associate professor), one pediatrician (assistant professor) and one neonatologist. Subsequently, pain assessment of all infants was performed by the experts and their PIPP-R scores were determined. The adjusted total item correlations obtained using a two-way mixed effect model were used to weighted the observers. The observer with the greatest total item correlations was determined, and the validity and reliability analyses were performed on the evaluations of this observer.

Ethical aspects

Permission was obtained from the author who developed the scale before study commencement. Necessary permissions were then obtained from the related institution and the Ethics Committee of the Erciyes University Faculty of Medicine. Before recording using the video camera, the parents of the newborn infants were informed of the study, that the data would be used for this study only, and that confidentiality principles would be followed. They signed an informed consent form.

Results

Descriptive characteristics of infants

In the present study, it was determined that 56% of the infants were boys, 30% were in the gestational age range of 26–31 weeks, 35.5% were in the gestational age range of 32–36 weeks and 34.5% were in the gestational age range of 37–42 weeks, 46% of the infants were at the natal age of 3 days. The birth weight of 78% of infants was >1.501 g, and 74.5% were born by cesarean section (Table 1).

Content validity for the PIPP-R

The CVI and CVR were calculated according to Lawshe's technique and the values are shown in Table 2. The CVR values of the scale items were 0.8 for change in heart rate, 0.8 for a decrease in oxygen saturation, 0.8 for brow bulge, 1.0 for eye squeeze, 0.8 for nasolabial furrow, 1.0 for baseline behavioral state and 1.0 for gestational age, the CVI was 0.88 for these seven items.

Construct validity for the PIPP-R

The factor analysis results of the PIPP-R are shown in Table 3. In this study, the seven items of the PIPP-R comprised three components (subgroup) with Eigenvalues >1.000. The first component explained 36.40% of the total variance, the second component explained 23.58% of the total variance and the third component explained 22.89% of the total variance. All three components explained 82.87% of the total variance. The varimax rotation method was used to determine the scale items forming the components and factor loads.

The factor loads obtained according to the principal component analysis and the varimax rotation methods are shown in Table 4. In this study, items with factor loads of >0.70 were accepted as significant for the factor structure. Accordingly, it was determined that the first factor was composed of the items of brow bulge, eye squeeze, and nasolabial furrow; the second factor included the change in heart rate and the decrease in oxygen saturation, and the third factor comprised the gestational age and baseline behavioral state items.

Table 5
Item total correlation values of the PIPP-R items.

Scale items	$\bar{x} \pm ss$	Med (Min-max)	Corrected item total correlation values	Cronbach's alpha values when the item was deleted
Change in heart rate	1.98 ± 0.75	2(0–3)	0.514	0.830
Decrease in oxygen saturation	2.09 ± 1.10	3(0–3)	0.553	0.832
Brow bulge	2.32 ± 0.67	2(0–3)	0.686	0.809
Eye squeeze	2.09 ± 0.76	2(0–3)	0.734	0.799
Naso-labial furrow	2.03 ± 0.87	2(0–3)	0.731	0.796
Gestastional age	0.96 ± 0.80	1(0–3)	0.464	0.837
Baseline behavioral state	1.72 ± 0.91	2(0–3)	0.565	0.823

Internal consistency coefficient (Cronbach's alpha coefficient) and item total correlations

In this study, the internal consistency coefficient (Cronbach's alpha coefficient) of the PIPP-R was 0.840 and the scale was determined to be highly reliable. When the items of the scale were omitted, the Cronbach's alpha values were 0.796–0.837 (Table 5), and it was decided that no items should be omitted from the scale.

The adjusted total item correlation values for the PIPP-R items are also shown in Table 5. It was found that the adjusted total item correlations were 0.464–0.734. As the adjusted total item correlations of the items were >0.40, it was decided that no items should be omitted from the scale.

Inter-observer consistency

The intraclass correlation coefficients (ICC) values were used to evaluate the agreement among the three observers. The ICC values were 0.944–1.000 (Table 6). The agreement between the three observers, who were specialized in their fields, was evaluated as very good.

Selectivity and sensitivity results for PIPP-R

In the calculations for the formulas used in the selectivity and sensitivity analysis of this study, it was found that the sensitivity of the PIPP-R was 91.5% and its selectivity was 88.5% (Tables 7 and 8). Accordingly, the test produced 0.09 incorrect results for seven correct results.

Table 9 shows the mean PIPP-R scores of the heel lance and diaper changing procedures of the newborns, in terms of the gestational age to determine the selectivity and sensitivity of the PIPP-R.

The PIPP-R total mean scores of the newborn infants in the gestational age range of 26–31 weeks, 32–36 weeks, and 37–42 weeks were determined as 16.30 ± 3.24, 13.69 ± 3.05, and 9.92 ± 3.80, respectively, during the heel lance procedure. A statistically significant difference was determined between the groups ($p < 0.001$).

The PIPP-R total mean scores of the newborn infants in the gestational age ranges of 26–31, 32–36, and 37–42 weeks were determined as 11.91 ± 2.53, 8.12 ± 2.9, and 5.21 ± 2.87, respectively, during the diaper changing procedure. A statistically significant difference was determined among the groups ($p < 0.001$).

In comparing the PIPP-R pain scores between gestational age groups *via* advanced analysis (Tukey HSD test), it was determined that the highest scores were in the youngest infant group (26–31 weeks) for both the painful and the painless procedures ($p < 0.001$).

The mean painful and painless procedure PIPP-R scores were compared in terms of the gestational age. Accordingly, the PIPP-R total mean scores of this scale were 13.17 ± 4.24 for the heel lance procedure and 8.26 ± 3.86 for the diaper change procedure. The PIPP-R mean scores of the newborn infants in the gestational age ranges of 26–31, 32–36, and 37–42 weeks based on their gestational ages were 16.30 ± 3.24–11.91 ± 2.53, 13.69 ± 3.05–8.12 ± 2.90, and 9.92 ± 3.80–5.21 ± 2.87, respectively for the painful and painless procedures. Accordingly, it was determined that the mean PIPP-R scores obtained in each group during the diaper changing procedure were lower than the mean scores obtained during the heel lance procedure, and a statistically significant difference was found between the mean scores ($p < 0.001$).

Table 7
Grouping of scores obtained from the PIPP-R based on type of the procedure.

Assessment	Actual case			
	Heel lance (painful)		Diaper change (painless)	
	Number	Percentage	Number	Percentage
Painful	183 (A)	91.5	23 (B)	11.5
Painless	17 (C)	8.5	177 (D)	88.5
Total	200	100.0	200	100.0

Discussion

In the present study, language, content, and construct validity were assessed to determine the validity of the scale, and the internal consistency coefficient, item total correlations, and observer reliability were assessed to determine the reliability of the scale during the validity and reliability stages of the PIPP-R. The scale was applied in extreme groups (different gestational age; painful and painless procedure) during the selectivity and sensitivity stage.

The factor analysis method was used in the present study to determine the construct validity of the seven item PIPP-R. When the factor structure of the scale was examined, it was found to include three factors; the first factor comprised the brow bulge, eye squeeze, and nasolabial furrow; the second factor comprised the change in heart rate and decrease in oxygen saturation; and the third factor comprised the initial behavioral state and gestational age. Stevens et al. also reported that the PIPP-R consisted of three factors (Stevens et al., 2014).

In the present study, the Cronbach's alpha coefficient was used to determine the internal consistency of the PIPP-R, which was found to be highly reliable. In a study conducted by Sadeghi et al., Cronbach's alpha value of the PIPP-R was determined as 0.71 for the painful procedure (Sadeghi et al., 2017).

In the present study, the observer reliability method was used to determine the reliability of the PIPP-R. The ICC values were used to evaluate the agreement between three observers, who were specialized in their fields, and were found to be 0.944–1.000 with the agreement between the observers evaluated as very good. In a previous study, Sadeghi et al. found that the ICC coefficient values were 0.99 for the painful procedure (Sadeghi et al., 2017). In the present study, it was found study that the sensitivity–of the PIPP-R was 91.5%, whereas its selectivity was 88.5%. Jonsdottir and Kristjansdottir stated that determining the selectivity of a scale requires application of the scale at extreme groups and Stevens et al. stated that this one of the most effective methods for determining internal consistency (Jonsdottir & Kristjansdottir, 2005; Stevens et al., 1996; Stevens et al., 2014). Therefore, they evaluated the PIPP-R at different gestational age groups and on the painful and painless procedures in their study (Gibbins et al., 2014; Stevens et al., 2014). In this regard, the process of applying the

Table 6
Evaluation of agreement between three observers.

Scale items	Intra-class correlation coefficients values	
	ICC	p
Change in heart rate	0.963	<0.001
Decrease in oxygen saturation	0.969	<0.001
Brow bulge	0.944	<0.001
Eye squeeze	0.950	<0.001
Naso-labial furrow	0.956	<0.001
Gestastional age	1.000	–
Baseline behavioral state	0.980	<0.001

Table 8
Sensitivity and selectivity values of PIPP-R.

Variables	Sensitivity and selectivity formulas
Sensitivity	A/(A + C) =183/200 = 0.915
Selectivity	D/(B + D) =177/200 = 0.885
Positive predictive value	A/(A + B) = 183/(23 + 183) = 0.88
Negative predictive value	(D/(C + D) = 177/(17 + 177) = 0.91
Positive likelihood ratio	Sensitivity/(1−Selectivity) = 0.915/(1–0.885) = 7.956
Negative likelihood ratio	(1−Sensitivity)/Selectivity = (1–0.915)/0.885 = 0.096

Table 9
PIPP-R mean scores of heel lance and diaper change procedure of infants based on gestational ages.

Gestational age	Heel lance		Diaper change		Test	p
	$\bar{x} \pm ss$	Med (Min-Max)	$\bar{x} \pm ss$	Med (Min-Max)		
26–31 weeks	16.30 ± 3.24	(7–20)	11.91 ± 2.53	(5–18)	$t = 8.579$	p = 0.000
32–36 weeks	13.69 ± 3.05	(5–18)	8.12 ± 2.90	(3−13)	$t = 11.650$	p = 0.000
37–42 weeks	9.92 ± 3.80	(2–16)	5.21 ± 2.87	(1−11)	$t = 8.729$	p = 0.000
Total score	13.17 ± 4.24	(2−20)	8.26 ± 3.86	(1–18)	$t = -16.637$	p = 0.000
Test	F = 58.114 p = 0.000		F = 92.601 p = 0.000			

scale in the extreme groups was used in the present study to determine the selectivity and sensitivity of the PIPP-R. For this purpose, the mean scores of the newborns in different gestational weeks (26–31, 32–36, and 37–42 gestational age range) were compared in the PIPP-R assessments for painful (heel lance) and painless (diaper change) procedures. Stevens et al. stated that the pain experienced by a newborn infant was mild if the mean scores of the PIPP-R were 0–6 points, moderate at 7–12 points, and severe at 13–21 points (Stevens et al., 2014).

In the present study, the PIPP-R total mean scores for painful and painless procedures were determined to be 13.17 ± 4.24 for the heel lance procedure and 8.26 ± 3.86 for the diaper change procedure. As the gestational age decreased, the pain perceived by the newborns was measured as severe, and there was a difference between the mean scores of the painful and painless procedure groups. Various factors affect the perception of pain in newborn infants. It is reported that there is a correlation between the gestational age, which is one of these factors, and the perception of pain; as gestational age decreases, the severity of the perceived pain increases (Valeri & Martins Linhares, 2012; Walker, 2017). This situation has been interpreted as PIPP-R is distinguishing between the painful and painless procedures, thus, it is sensitive and selective. Similarly, Gibbins et al. used the PIPP-R at the extreme groups in their study and determined that there was a statistical difference between the mean PIPP-R scores and the total mean scores of the painful and painless procedures in terms of the gestational age ($p < 0.05$) (Gibbins et al., 2014). Sadeghi et al. also indicated that the scores calculated by the PIPP-R during diaper change were lower than those calculated during a heel prick procedure. They determined a statistically significant difference between the mean PIPP-R scores of newborn infants for painful and painless procedures (Sadeghi et al., 2017).

Implications for nursing practice

The results of the present study were consistent with the results from the analysis of the original version of the scale. It was determined that PIPP-R is a valid, reliable, selective, and sensitive scale that can be used in the pain assessment of infants. The diagnosis and management of pain require teamwork. The nurses included in this team must not only assess pain in the painful and needle-based procedures in infants, but must also measure the efficiency of non-pharmacological nursing interventions to reduce pain by using the PIPP-R (Akcan & Polat, 2017; Eroğlu & Arslan, 2018).

Limitations

The present study aimed to determine the reliability, selectivity, and sensitivity of the PIPP-R scale, therefore, the effect of the applied procedure on the pain levels of newborns and the factors affecting pain were not considered. The scale was developed for preterm and term newborns and can only be used for these age groups.

Conclusion

PIPP-R is a multidimensional pain assessment tool that measures physiological and behavioral parameters together. In this study, the reliability of the scale was measured using Cronbach's alpha coefficient and observer reliability, and the scale was found to be highly reliable. The validity of the scale was evaluated by factor analysis, which revealed that the scale was composed of three physiological, behavioral and contextual factors. In addition, the scale is considered selective in measuring the pain of different-age-gestated newborns and sensitive in distinguishing between painful and painless procedures.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Authorship statement

All listed authors meet the authorship criteria and that all authors are in agreement with the content of the manuscript.

Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank Ferhan Elmalı (Associate Professor, PhD, in Department of Biostatistics and Medical Informatics, Faculty of Medicine, Izmir Kâtip Çelebi University) for statistical analysis.

References

Aarnoudse-Moens, C. S., Weisglas-Kuperus, N., van Goudoever, J. B., & Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*, *124*, 717–728. https://doi.org/10.1542/pads.2008-2816.

Akcan, E., & Polat, S. (2017). Pain in newborns and the nurse's role in pain management. *Acibadem University Health Sciences Journal*, *2*, 64–69 (in Turkish).

Akcan, E., & Yiğit, R. (2015). Turkish reliability and validity study of premature ınfant pain profile. *Fırat University Medical Journal of Health Sciences*, *29*, 97–102 (in Turkish).

Alpar, R. (2016). *Applied statistics and validity-reliability.* Detail Publishing, 508–520 (In Turkish).

Brummelte, S., Grunau, R. E., Chau, V., et al. (2012). Procedural pain and brain development in premature newborns. *Annals of Neurology*, *71*, 385–396. https://doi.org/10.1002/Ana. 22267.

Büyüköztürk, Ş. (2015). *Handbook of data analysis for social sciences* (Extended 21st ed.). Ankara: Pegem Academy, 179–194 (In Turkish).

Çapık, C. (2014). The use of confirmatory factor analysis of reliability and validity. *Anatolian Journal of Nursing and Health Sciences*, *17*(3), 196–205 (in Turkish).

Carbajal, R., Rousset, A., Danan, C., et al. (2008). Epidemiology and treatment of painful procedures in neonates in intensive care units. *JAMA*, *300*, 60–70. https://doi.org/10.1001/Jama.300.1.60.

Carter, B. S., & Brunkhorst, J. (2017). Neonatal pain management. *Seminars in Perinatology*, *41*, 111–116. https://doi.org/10.1053/j.semperi.2016.11.001.

Dolgun, G., & Bozlak, Ş. (2017). The effect of nonpharmacologic pain controls during examination for retinopathy of prematurity. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, *46*(5), 709–715. https://doi.org/10.1016/j.John.2017.06.008.

Eroğlu, A., & Arslan, S. (2018). The perception of the neonatal pain assessment and management. Duzce University Journal of the Institute of. *Health Sciences*, *8*(1), 52–60 (in Turkish).

Gibbins, S., Stevens, B. J., Yamada, J., et al. (2014). Validation of the premature ınfant pain profile-revised (PIPP-R). *Early Human Development*, *90*, 189–193. https://doi.org/10.1016/j.earlhumdev.2014.01.005.

Hall, R. W., & Anand, K. J. S. (2005). Physiology of pain and stress in the newborn. *NeoReviews*, *6*, 61–67. https://doi.org/10.1542/neo.6-2-e61.

Jonsdottir, R. B., & 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*, 598–605.

Kabataş, E. U., Dursun, A., Beken, S., Dilli, D., Zenciroğlu, A., & Okumuş, N. (2016). Efficacy of single dose oral paracetamol in reducing pain during examination for Retinopathy of Prematurity: A Blinded Randomized Controlled Trial. *Indian Journal of Pediatrics*, *83* (1), 22–26. https://doi.org/10.1007/s12098-015-1765-8.

Karakoç, F. Y., & Dönmez, L. (2014). Basic principles in scale development studies. *Medical Education World Journal*, *40*, 39–49 (in Turkish).

Küçük Alemdar, D., & Kardaş Özdemir, F. (2017). Effects of having preterm ınfants smell amniotic fluid, mother's milk, and mother's odor during heel stick procedure on pain, physiological parameters, and crying duration. *Breastfeeding Medicine*, *12*, 297–304. https://doi.org/10.1089/bfm.2017.0006.

Roofthooft, D. W. E., Simons, S. H. P., Anand, K. J. S., Tibboel, D., & van Dijk, M. (2014). Eight years later, are we still hurting newborn infants? *Neonatology*, *105*, 218–226. https://doi.org/10.1159/000357207.

Sadeghi, A., Rassouli, M., Gharehdaghi, F. A., Nourian, M., Nasiri, M., Farhadi, R., & Farahan, A. S. (2017). Validation of the Persian version of premature ınfant pain profile-revised in hospitalized ınfants at the neonatal ıntensive care units. *Iranian Journal of Pediatrics*, *27*(5), e10056. https://doi.org/10.5812/ijp.10056.

Şener Taplak, A., & Erdem, E. (2017). A comparison of breast milk and sucrose in reducing neonatal pain during eye exam for retinopathy of prematurity. *Breastfeeding Medicine*, *12*, 305–310. https://doi.org/10.1089/bfm.2016.0122.

Stevens, B., Johnston, C., Petryshen, P., & Taddio, A. (1996). Premature ınfant pain profile: Development and initial validation. *The Clinical Journal of Pain*, *12*, 13–22 (8722730).

Stevens, B., McGrath, P., Gibbins, S., et al. (2003). Procedural pain in newborns at risk for neurologic impairment. *Pain*, *105*, 27–35. https://doi.org/10.1016/S0304-3959(03)00136-2.

Stevens, B. J., Gibbins, S., Yamada, J., et al. (2014). The premature infant pain profile-revised (PIPP-R): Initial validation and feasibility. *The Clinical Journal of Pain*, *30*, 238–243. https://doi.org/10.1097/AJP.0b013e3182906aed.

Valeri, B. O., & Martins Linhares, M. B. (2012). Pain in preterm infants: Effects of sex, gestational age, and neonatal illness severity. *Psychology & Neuroscience*, *5*, 11–19. https://doi.org/10.3922/j.psns.2012.1.03.

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*, 334–344.

Walker, S. M. (2017). Translational studies identify long-term impact of prior neonatal pain experience. *Biennial Review of Pain*, *0*, 1–14. https://doi.org/10.1097/j.pain.0000000000000784.

Yurdugül, H. (2005). The use of content validity index for the content validity of the scale development. *National Educational Sciences Congress, 28–30 September* (pp. 1–6). Pamukkale University Faculty of Educational Sciences XIV (in Turkish).