INTRODUCTION
Preterm and term newborns experience pain and stress due to numerous and very different reasons such as intubation, venipuncture, and nasogastric/orogastric tube insertion in Neonatal Intensive Care Units (NICUs)\(^1\). Newborns experience approximately 70 stressful procedures\(^2\) or 51 painful stimuli per day in the NICU\(^3\). Newborns may respond to pain in an exaggerated or attenuated manner as a result of frequent painful and invasive procedures\(^4\). The pain experienced by the newborns can not only prevent his/her behaviors, family infant interaction, and infant’s adaptation to the outside world, but also cause changes in the development of brain and senses and affect the growth negatively\(^1\).

Failure to appropriately assess the pain of newborns may result in delayed treatment and negative consequences\(^1\).

Emotional status that cannot be expressed verbally by infants is involved in pain assessment, thus resulting in problems about the definition and treatment of pain. Newborns are dependent on others for the definition and treatment of pain\(^1\). Since newborns cannot express themselves, they show pain in behavioral and physiological ways\(^5\). Physiological changes related to pain in newborns include the changes in heart rate, respiratory rate, blood pressure, and oxygen and carbon dioxide levels in the blood. Behavioral changes related to pain include crying, facial expressions, motor movements, and behavioral status changes\(^7\).

For the objective assessment of pain, the American Academy of Pediatrics recommends the use of validated and reliable scales\(^8\). Many scales have been developed for the measurement and evaluation of pain in newborns\(^9\). It has been reported that pain scales that are not suitable for the unsuitable population are used,

\(^1\)Yuksek Ihtisas University, Faculty of Health Sciences, Department of Pediatric Nursing – Ankara, Turkey.
\(^2\)Bagcılar Training and Research Hospital, Neonatal Intensive Care Unit – Istanbul, Turkey.
\(^3\)Istanbul University-Cerrahpasa, Florence Nightingale Faculty of Nursing, Department of Pediatric Nursing – Istanbul, Turkey.
Corresponding author: bdonmez@istanbul.edu.tr
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on December 05, 2020. Accepted on December 13, 2020.
so it is important to use appropriate pain scales for newborns. Hand et al. developed COVERS and reported that it is a valid scale for both preterm and term newborns. O’Sullivan et al. proved that the scale is valid and reliable. Considering the negative consequences of acute pain in newborns, it is very important to evaluate the pain of the newborn with a valid and reliable scale in daily clinical practice in the NICU. This study was conducted to determine the reliability and validity of COVERS, which has not been used yet in newborns in Turkey.

METHODS

Setting, design, and sample
This study was conducted using a methodological design in a 16-bed NICU of a public hospital between January and June 2018. The case number was determined for a moderate effect size \( [(\Delta: 0.50), \beta: 0.20 (80\% \text{ power: } 1-\beta) \text{ and } \alpha: 0.05] \) was found as 28 newborns. Forty-one newborns who met the criteria were included in this study. The inclusion criteria for the infants were determined as follows: being 24 weeks old and older, receiving no analgesic treatment within the last 12 h, and having no congenital anomaly.

Newborn information form
The form included questions about gestational week, weight, length of hospital stay (in days), gender, diagnosis, respiratory support, and the type of receiving respiratory support.

Neonatal infant pain scale
The scale consisting of behavioral parameters was developed by Lawrence et al. It consists of six subscales such as facial expression, crying, breathing patterns, arm movement, leg movement, and state of arousal. The scores of the scale range between 0 and 7, and the score higher than 4 points signifies pain. The Cronbach’s \( \alpha \) value of neonatal infant pain scale (NIPS) was reported as 0.95, 0.87, and 0.88 before, during, and after the procedure, respectively. Akdovan adapted the scale into Turkish, and the Cronbach’s \( \alpha \) value of the scale was found to be between 0.83 and 0.86. In this study, the Cronbach’s \( \alpha \) values were 0.79 and 0.87 during and after heel lance procedure, respectively.

Preterm infant pain profile
The scale was developed for 28- to 36-week preterm infants by Stevens et al. It evaluates behavioral (e.g., eyebrows, eyes, nasolabial furrow, and facial movement) and physiological (e.g., heart rate and oxygen saturation) parameters of the infant. The score of the scale ranges between 0 and 21. While a score of <5 indicates no pain, a score of >10 indicates moderate to severe pain. The Cronbach’s \( \alpha \) value of preterm infant pain profile (PIPP) was found to be between 0.59 and 0.76. The reliability and validity study of the scale was conducted by Alkan and Yiğit in Turkey. The Cronbach’s \( \alpha \) value of the scale was between 0.68 and 0.78. In this study, the Cronbach’s \( \alpha \) values were 0.72 and 0.32 during and after heel lance procedure, respectively.

COVERS pain scale
The scale, which was developed by Hand et al., consists of six subscales and depends on physiological (e.g., oxygen requirement and vital signs) and behavioral (e.g., crying, expression, resting, and signaling distress) measurements (Table 1). The subscales of the COVERS scale are scored with 0, 1, and 2 points, and the scale score ranges between 0 and 12. The validity of the scale in 27- to 40-week infants was analyzed. The scale was valid and reliable in newborns older than 24 weeks, and it can be applied in the clinic.

Forward and backward translation
The researchers translated COVERS scale into Turkish, and the translated scale was checked by 10 pediatric nursing academicians. The scale translated into Turkish was translated back into English by a linguist. The original scale and the translated English scale were checked and found similar by another linguist. The language validity was performed in this way.

Content validity
The scale was evaluated by 10 pediatric nursing academicians. Each expert was asked to evaluate the relevance level of each item about the purpose of the questionnaire with a 4-point Likert scale as “completely appropriate” to “not appropriate.” The content validity index of the scale was calculated as 0.91.

Study protocol
This study was conducted in supine position during both diaper change and heel lance procedures between 10:00 a.m. and 11:00 a.m. during weekdays while the infants were awake, and the process was recorded on video. Both diaper change and heel lance procedures were applied to each infant. The scales were applied 1 min before, during, 1 min after the diaper change, and during heel lance procedure (i.e., at the first time when manual lancet penetrated) and 1 min after heel lance procedure (i.e., applying cotton). The application lasted for an average of 5 min.

Concurrent validity and construct validity
For concurrent validity, the infants were evaluated with the scales such as COVERS, PIPP, and NIPS. For construct validity, they
The validity and reliability of COVERS were assessed using the scales in painful and nonpainful procedures. According to the literature, a nonpainful procedure was determined as the diaper change, and the painful procedure was determined as heel lance procedure.

**Inter-rater reliability**

The diaper change and heel lance procedures were performed by a nurse included in this study. Another nurse involved in this study recorded the application process and the monitor showing heart rate and oxygen saturation of the infant via a video recorder. The nurse making the video recording in this study and the academician nurse included in this study watched the videos of the infants independently and evaluated the infants according to COVERS, PIPP, and NIPS.

**Ethics statement**

Ethics committee approval permission (IRB: 2017,12,4,05,021) was obtained from the hospital. Permissions to use scales were obtained by authors via e-mail. The written informed consent was obtained from the parents.

**Statistical analysis**

The IBM SPSS Statistics 22 (IBM SPSS, Turkey) was used for the statistical analyses. Normality assessment of the variables was made by using the Shapiro-Wilk test. The statistical significance level was set at 0.05. The internal consistency of COVERS at each time point was established using mean inter-item correlations, corrected item-total correlations, and the Cronbach’s α reliability coefficients. Inter-rater reliability was established using kappa measure of agreement for categorical data and intraclass correlation coefficients (ICCs) for the continuous data. While the Spearman’s rho correlation coefficient was used for concurrent validity, the Wilcoxon signed rank test was used for construct validity.

**RESULTS**

The characteristics of the infants were shown in Table 2.

**Concurrent validity**

In <37 weeks (n=28), a statistically significant correlation was determined between COVERS and PIPP scores during (r=0.768, p<0.001) and after (r=0.617, p<0.001) heel lance procedure, and between COVERS and NIPS scores during (r=0.785, p<0.001) and after (r=0.800, p<0.001) heel lance procedure.

In >37 weeks (n=13), a statistically significant correlation was determined between COVERS and PIPP scores during (r=0.854, p<0.001) and after (r=0.869, p<0.001) heel lance procedure.
and between COVERS and NIPS scores during (p=0.823, 
p<0.001) and after (p=0.951, p<0.001) heel lance procedure.

Construct validity
Table 3 shows the distribution of the COVERS mean scores 
of the newborns before, during, and after the diaper change 
and heel lance procedures.

Internal consistency
The item-total correlation values of COVERS were 0.32– 
0.82 and 0.39–0.86 during and after heel lance procedure. 
The Cronbach’s α values of COVERS were 0.77 and 0.83 
during and after heel lance procedure.

Inter-rater reliability
Table 4 shows the kappa results during diaper change and heel 
lance procedures. ICC values obtained for COVERS total score 
were 0.741 (95%CI, 0.514–0.862) during diaper change procedure 
and 0.579 (95%CI, 0.211–0.776) during heel lance procedure.

DISCUSSION
In a systematic review, it was stated that the validity and 
reliability of a scale must necessarily have construct validity, 
internal consistency, and inter-rater reliability¹⁷. In this study, 
content validity, concurrent validity, and construct validity 
were tested for the validity of the COVERS scale. For the 
reliability of the scale, internal consistency and inter-rater 
reliability were tested.

For correlation values, it was reported that values between 
0.70 and 0.89 showed a high correlation and values between 0.90 
and 1.00 showed a very high correlation¹⁸. It was determined in 
a study conducted by Hand et al.¹⁰ that while COVERS showed 
a high degree of correlation with PIPP (r=0.84) in preterm 
infants, it showed a very high degree of correlation with NIPS 
(r=0.95) in full-term infants. In this study, COVERS was found 
to be highly correlated with PIPP and NIPS. According to the 
results of this study, COVERS was found to have concurrent 
validity. Construct validity is defined as the degree to which 
a test measures what it is supposed to measure¹⁹. It was found 
in this study that there was a significant difference between 
the mean scores of COVERS in nonpainful and painful pro-
cedures. The results of this study are consistent with the liter-
ature²⁰. In this study, it was determined that COVERS, which 
was adapted to Turkish, had construct validity.

In the literature, it has been reported that 0.30 and above is 
accepted as the optimum for corrected item-total correlation, 
and it becomes perfect as it approaches to 1²⁰. In this study,
since item-total correlation of COVERS was above 0.30 during and after heel lance procedure and showed that the items were appropriate, no item was excluded from the scale. It was reported in the study by O’Sullivan et al.\textsuperscript{11} that corrected item-total correlation of COVERS was 0.19–0.68 during heel lance procedure and only the score of the item “oxygen requirement” was below 0.30. In contrast to the literature\textsuperscript{11}, the score of this item was found to be above 0.30 in this study.

The Cronbach’s $\alpha$ value between 0.70 and 0.95 is reported to be an acceptable value\textsuperscript{21,22}. In the literature\textsuperscript{11}, the internal consistency of COVERS was 0.78. The results of this study were found to be compatible with the literature. For kappa statistic values, it is reported that <0.20 is weak agreement, 0.20–0.40 is acceptable agreement, 0.40–0.60 is moderate agreement, 0.60–0.80 is good agreement, and 0.80–1.00 is perfect agreement\textsuperscript{23}. In the study by O’Sullivan et al.\textsuperscript{11}, IC\textsuperscript{Cs} for scores of COVERS were 0.82 (95%CI, 0.72–0.88) at baseline and 0.80 (95%CI, 0.69–0.87) during heel lance. In this study, it was found that there was a moderate correlation in IC\textsuperscript{Cs} for scores of COVERS during both diaper change and heel lance procedures.

It has been reported in the literature that PIPP and NIPS are frequently used to evaluate acute pain of preterm and term newborns\textsuperscript{9}. The COVERS scale includes the parameters of both the NIPS and PIPP scale. However, the criteria used for scoring on the COVERS scale include newborns with a wider week\textsuperscript{10}. Although crying is a parameter of behavioral responses to pain, it is unlikely that an intubated infant will have a high-pitched crying\textsuperscript{10}. Compared with NIPS\textsuperscript{13}, the COVERS scale included visible crying within the behavioral parameter\textsuperscript{10}. Compared with the PIPP scale, the COVERS scale has brought a new perspective to oxygen requirements. Since the oxygen requirement of the infant is not always an indicator of pain, the COVERS scale focuses on the change of oxygen demand rather than its value\textsuperscript{10}.

Table 3. Distribution of COVERS pain mean scores before, during, and after the diaper change and heel lance procedure (n=41) determined by using Wilcoxon signed rank test.

<table>
<thead>
<tr>
<th>COVERS</th>
<th>Before</th>
<th>During</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diaper change</td>
<td>0.636</td>
<td>3.415</td>
<td>1.439</td>
</tr>
<tr>
<td>Heel lance</td>
<td>1.341</td>
<td>5.122</td>
<td>1.829</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
<td>0.003</td>
<td>0.439</td>
</tr>
</tbody>
</table>

Table 4. COVERS kappa measure of agreement analysis results of two observers during diaper change and heel lance procedure (n=41).

<table>
<thead>
<tr>
<th>COVERS</th>
<th>During diaper change</th>
<th>During heel lance procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa</td>
<td>SE</td>
</tr>
<tr>
<td>Crying</td>
<td>0.681</td>
<td>0.134</td>
</tr>
<tr>
<td>Oxygen requirement</td>
<td>0.448</td>
<td>0.152</td>
</tr>
<tr>
<td>Vital signs</td>
<td>0.394</td>
<td>0.139</td>
</tr>
<tr>
<td>Expression</td>
<td>0.561</td>
<td>0.121</td>
</tr>
<tr>
<td>Resting</td>
<td>0.345</td>
<td>0.150</td>
</tr>
<tr>
<td>Signaling distress</td>
<td>0.473</td>
<td>0.155</td>
</tr>
</tbody>
</table>

SE: standard error; CI: confidence interval.
CONCLUSIONS
It was concluded that based on content validity, concurrent validity, and construct validity analyses of COVERS adapted into Turkish, it is a valid scale, and based on internal consistency and inter-rater reliability analyses, it is a reliable scale. COVERS can be used in all newborns with a gestational age of ≥27 weeks. In future studies, it may be suggested to adapt the COVERS scale to different cultures and to investigate the validity and reliability of the COVERS scale on infants in the postneonatal period.

ACKNOWLEDGMENTS
The authors thank the preterm infants’ families of the Neonatal Intensive Care Units where the study was conducted.

AUTHORS’ CONTRIBUTIONS
MCI: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.
NUÖ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Validation, Writing – original draft.
BM: Conceptualization, Funding acquisition, Methodology, Validation, Writing – original draft, Writing – review & editing.
EC: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Software, Validation, Writing – original draft.

REFERENCES
The validity and reliability of COVERS


