Interrater Reliability and Concurrent Validity of the Neonatal Eating Outcome Assessment

Roberta Pineda, Lara Liszka, Jenny Kwon, Michael Wallendorf

Importance: Few neonatal feeding assessments are currently available, and the Neonatal Eating Outcome Assessment is the only one that identifies feeding impairment while considering the developmental changes that occur from preterm birth to term-equivalent age.

Objective: To determine the interrater reliability and concurrent validity of the Neonatal Eating Outcome Assessment.

Design: Prospective, observational study.

Setting: Level 4 neonatal intensive care unit.

Participants: A convenience sample of 7 neonatal therapists participated in reliability testing. For concurrent validity, a prospective cohort of 52 preterm infants born ≤ 32 wk gestation had feeding assessed at term-equivalent age.

Outcomes and Measures: Intraclass correlations (ICCs) and Fleiss's κ statistics were used to define reliability across therapists, who independently scored five videos of preterm infants orally feeding using the Neonatal Eating Outcome Assessment. Concurrent validity was determined by evaluating relationships between the Neonatal Oral Motor Assessment Scale (NOMAS) and the Neonatal Eating Outcome Assessment using an independent-samples t test and χ^2 analysis.

Results: The ICC for the Neonatal Eating Outcome Assessment total score was 0.90 (confidence interval [CI] [0.70, 0.99]). Fleiss's κ scores for the 19 scorable items on the Neonatal Eating Outcome Assessment had predominately moderate, fair, and slight agreement, with 3 items having poor agreement. Dysfunctional NOMAS scores were related to lower Neonatal Eating Outcome Assessment scores ($t[49.4] = 3.72$, mean difference = 12.2, 95% CI [5.60, 18.75], $p = .001$).

Conclusions and Relevance: The Neonatal Eating Outcome Assessment has excellent reliability. Concurrent validity was established.

What This Article Adds: This article reports that the final version of the Neonatal Eating Outcome Assessment (Version 5.7) has excellent interrater and concurrent validity and is an important tool to assess the occupation of infant feeding in clinical practice.

Feeding challenges are common in preterm infants, and 1 in 5 preterm infants demonstrate feeding problems (de Kieviet et al., 2009; Dodrill, 2011; Gennattasio et al., 2015). Oral feeding is one of the final milestones achieved before discharge from the neonatal intensive care unit (NICU), and it is one of the primary reasons for therapy consultation in the NICU (de Kieviet et al., 2009).

Oral feeding attempts typically begin around 33–34 wk postmenstrual age (PMA) for the very preterm infant but usually no earlier than 32 wk PMA (Barlow, 2009; Bertoncelli et al., 2012). During early oral feeding attempts, infants may demonstrate problems in the feeding process, which may or may not be related to later success with oral feedings. Feeding problems related to immaturity can result from a lack of reflexes that support feeding, such as rooting and sucking, or poor arousal (Einarsson-Backes et al., 1994; McGrath & Braescu, 2004). This transient period of feeding problems, which may be appropriate for the infant's level of immaturity, often resolves as the infant matures and approaches term-equivalent age (Gewolb et al., 2001; Ross & Browne, 2002). As a result, infants born preterm who initiate oral feeding between 32 and 40 wk PMA often struggle with oral feeding but are on a pathway toward normal

feeding once they mature (Ross & Browne, 2002). However, infants can also have feeding problems that can lead to long-term issues that need to be differentiated from problems related to immaturity. Tools that can differentiate problems in feeding in the context of PMA can enable clinicians to understand whether problems are transient or may have a long-term impact on feeding performance.

Few neonatal feeding assessments are currently available. Some, such as the Infant Breastfeeding Assessment Tool (Matthews, 1988), Mother–Baby Assessment (Mulford, 1992), LATCH (Jensen et al., 1994), Preterm Infant Breastfeeding Behavior Scale (Nyqvist et al., 1996), and Systematic Assessment of the Infant at Breast (Shrago & Bocar, 1990), are designed to assess breastfeeding. Others, such as the Early Feeding Skills assessment (Thoyre et al., 2005) and Infant-Driven Feeding Scales (Ludwig & Waitzman, 2007), are designed to assess feeding readiness. The Neonatal Oral Motor Assessment Scale (NOMAS; Palmer & Heyman, 1999) is used to identify problems with jaw and tongue patterns in neonates by assessing the first 2 min of an oral feeding by breast or bottle, and it can be used with both preterm and full-term infants (Palmer & Heyman, 1999). However, the NOMAS does not fully consider the developmental changes in oral feeding performance that occur after preterm birth. Only one tool, the Neonatal Eating Outcome Assessment, aims to identify problematic feeding behaviors while considering the developmental changes that occur from preterm birth to term-equivalent age (Pineda et al., 2018).

The Neonatal Eating Outcome Assessment is a developmental feeding assessment that can be used with preterm infants and full-term infants, who are breast or bottle fed, as soon as oral feeding is initiated through 4–6 wk corrected age. In a previous report, content validity was established using neonatal feeding experts, and the total score of the assessment had good to excellent reliability (Pineda et al., 2018). However, reliability statistics on each item of the assessment provided opportunities for optimization, leading to a final Version 5.7. The aim of this study is to determine the interrater reliability and concurrent validity of the Neonatal Eating Outcome Assessment (Version 5.7).

Method

This study was approved by the Washington University Human Research Protection Office and consisted of two parts. First, interrater reliability was evaluated using 7 neonatal therapists as participants to assess five videos of oral feeding so that scoring could be compared across raters. A waiver of consent was granted for the neonatal therapist participants. Second, concurrent validity was evaluated using a prospective longitudinal cohort of 52 preterm infants born ≤ 32 wk gestation whose parents signed informed consent. Feeding was assessed at term-equivalent age using the Neonatal Eating Outcome Assessment and the NOMAS, and scores were compared.

Measures

The Neonatal Eating Outcome Assessment was developed to assess age-appropriate feeding skills, based on PMA. It has 19 scorable items (Figure 1), including 1 item that is scored to define readiness for oral feeding but is not included in the total assessment score. This assessment can be used with infants from the initiation of breast- or bottle-feeding through 4–6 wk corrected age. It can be used to differentiate normal versus abnormal feeding performance and can track feeding progress. Each item is given a letter score (see Figure 1), which is converted to a numerical score on the basis of the infant's PMA at the time of testing, using a rubric. The item-level scores are then tallied for a total Neonatal Eating Outcome Assessment score, which ranges from 18 to 90; higher scores indicate better feeding performance. The total score can be converted to a categorical score (*normal*, *questionable*, or *feeding challenged*), based on established ranges. Content validity has been established, and interrater reliability has been previously reported as good to excellent (Pineda et al., 2018).

The NOMAS is a 28-item observation checklist of infant tongue and jaw movement during the first 2 min of bottle-feeding or breastfeeding (Palmer et al., 1993). Developed by Marjorie Meyer Palmer, a speech-language pathologist and infant feeding specialist, the NOMAS is a widely used neonatal feeding assessment (da Costa et al., 2008; da Costa & van der Schans, 2008; Howe et al., 2007; Palmer et al., 1993; Zarem et al., 2013) that defines oral feeding

Figure 1. Neonatal Eating Outcome Assessment scorable items.

Item	Option
Section I: Prefeeding Behaviors	
I-1. Arousal/state organization	A. Unable to rouse to State 3, even with stimulation B. Short periods of arousal with stimulation C. Prolonged arousal sustained after stimulation D. Wakes spontaneously
I-2. Physiological stability	A. Poor physiological stability B. Intermittent periods of physiological stability with energy depletion C. Intermittent periods of physiological stability without energy depletion D. Physiological stability
I-3. Respiratory support	A. Ventilated B. Noninvasive respiratory support C. Minimal respiratory support D. Room air
I-4. Rooting and grasp	A. No response B. Weak or unilateral root and grasp C. Root and grasp with stimulation D. Spontaneous root and grasp E. Brisk, frantic response or avoidance
I-5. Initiation of sucking	A. None B. Requires moderate stimulation C. Requires minimal stimulation D. Normal E. Active avoidance
I-6. Tongue	A. Flaccid or nonresponsive B. Flat with some tongue cupping C. Elevated and retracted D. Tongue cupping with central groove E. Deviated to side, tongue thrusting, or tongue bunching
I-7. Nonnutritive sucking	A. Absent B. Arrhythmic C. Intermittent burst-pause pattern D. Normal burst-pause pattern E. Prolonged burst-pause pattern
Section II: Oral Feeding	
II-1. Suck–swallow–breathe coordination	A. Unable to coordinate sucking, swallowing, and breathing; immediate sequelae B. No independent suck–swallow–breathe sequencing; able to coordinate only when provided intervention C. Able to feed with intermittent interventions; independent swallow–breathe sequence D. Independently demonstrates appropriate suck–swallow–breathe sequence
II-2. Sucking burst length	A. No sucking burst B. Minimal sucks per burst C. Appropriate sucks per burst D. Long pattern of sucks per burst

performance as *normal*, *disorganized*, or *dysfunctional* (Braun & Palmer, 1985; Zarem et al., 2013). There are mixed reports on the reliability of the NOMAS (da Costa et al., 2016; Palmer et al., 1993; Zarem et al., 2013). Some concurrent validity, with relationships between NOMAS categorical scores and infant behavior and cerebral structure, has been reported (da Costa & van der Schans, 2008; Palmer et al., 1993; Zarem et al., 2013).

Interrater Reliability

Seven experts in neonatal therapy were recruited through convenience sampling. They were recruited by an email invitation to a network of colleagues and a list of therapists who had purchased the Neonatal Eating Outcome

Figure 1. Neonatal Eating Outcome Assessment scorable items. (Cont.)

II-3. Suction	A. No suction B. Minimal suction C. Normal D. Excessive suction
II-4. Oral tone	A. Flaccid B. Decreased tone C. Adequate for feeding D. Increased oral tone, clenched jaw, and/or pursed lips
II-5. Quality of sucking movements	A. Abnormal sucking pattern B. Moderate discoordination C. Mild or intermittent discoordination D. Mature sucking pattern
II-6. Behavioral response to feeding	A. Active avoidance or significant feeding stress responses B. Limited positive engagement and comfort during feeding C. Minimal to moderate positive engagement and comfort with feeding D. No discomfort
II-7. Fluid loss	A. Severe fluid loss B. Moderate fluid loss C. Minimal fluid loss D. No fluid loss
II-8. Swallow	A. High risk (clear, clinical indications of aspiration) B. Moderate aspiration risk C. Mild aspiration risk D. Normal swallow
II-9. Respiratory control during feeding	A. Respiratory compromise with markedly increased work of breathing B. Moderately increased work of breathing C. Mildly increased work of breathing D. Appropriate respiratory control during feeding
Section III: End of feeding	
III-1. Feeding completion	A. Stops due to physiological instability B. Stops due to fatigue (unable to continue) or feeder stopped the feeding C. Discomfort D. Completed feeding E. Completed feeding, but demonstrates discomfort after feeding
III-2. State maintenance	A. Reaches State 3 or 4 only briefly B. Maintains State 3 or 4 for 4–7 min C. Maintains State 3 or 4 for 8–10 min D. Maintains State 3–4 for >10 min E. Maintains high level of arousal (State 5 or 6) that interferes with feeding
III-3. Volume consumed	A. No milk consumption B. Minimal milk consumption C. Some milk consumption D. Half of feeding completed E. 75% of feeding consumed F. Full feeding

Assessment for use in their hospital. Participants were licensed and registered occupational therapists, physical therapists, or speech–language pathologists, who had practiced in the NICU for a minimum of 4 yr and were willing to devote a minimum of 2 hr to training, 5 hr to practice with the assessment, and 2 additional hr to reliability testing.

The therapists engaged in a 1-hr web-based training on Neonatal Eating Outcome Assessment administration and scoring. The webinar consisted of education on the purpose and use of the Neonatal Eating Outcome Assessment, assessment procedures, and scoring criteria. The therapists then engaged in an additional 1-hr education session via GotoMeeting (<https://www.gotomeeting.com/>), a web-based conferencing system that allowed them to join a meeting via their computer, laptop, tablet, or smartphone. All therapists joined the conference at the same time, along with the

author of the Neonatal Eating Outcome Assessment (first author Roberta Pineda). They observed videos of infant feeding, scored them, compared their scores with those of others, engaged in dialogue about what they were seeing, and had an opportunity to ask any questions.

After the 2 hr of training (1-hr web-based and 1-hr didactic training by means of online conferencing), therapists were instructed to practice using the Neonatal Eating Outcome Assessment by observing and scoring a minimum of 10 infants during oral feeding, before the reliability testing session approximately 3 wk later. These practice sessions ensured that the therapists were familiar with the scoring criteria and provided them with opportunities to observe multiple different types of infants.

The therapists reconvened on video conferencing via GotoMeeting to observe and score five videos using the Neonatal Eating Outcome Assessment. The videos were of preterm infants enrolled in an overarching study investigating the effect of positive sensory experiences on infant neurobehavior and feeding. For the overarching study, 100 preterm infants ≤ 32 wk gestation were enrolled from August 2017 to June 2018. Parents signed informed consent, and institutional review board approval was given to use the recorded videos for the current study. Infants were excluded if a congenital anomaly was confirmed or suspected, the parents did not speak English, or the infant was born >32 wk gestation.

For the videos, infants were fed according to standard NICU practices, after the physician gave approval, and feedings were advanced according to tolerance and success. Parents chose whether to bottle-feed or breastfeed. The full experience of an oral feeding, from before the nipple was placed in the mouth until the feeding was completed or stopped, was video-recorded between 35 and 44 wk PMA before NICU discharge. No instructions on how to feed the infant were given during the recording, which aimed to capture normal feeding experiences.

Five videos were randomly selected from the overarching study that included adequate visibility of the lateral aspect of the lips and jaw, involved the time before the nipple went into the infant's mouth, and were of adequate quality to enable scoring. Therapists were instructed to refrain from asking questions or making comments during the reliability session when videos were observed and scored to prevent biasing scores across raters. Therapists entered a letter score for each of the 19 scored items for each of the 5 infants into the Research Electronic Data Capture (REDCap; [Harris et al., 2009](#)) system by completing a REDCap survey. REDCap is a secure, web-based system for storing research data. Total scores were then calculated from the letter scores assigned to each item.

Concurrent Validity

To determine concurrent validity, a cohort of 52 preterm infants who were born ≤ 32 wk gestation, without congenital anomalies, and enrolled by the first week of life was used. Infants were part of two overarching studies ([Pineda et al., 2018, 2019](#)) in which infants had a full feeding video recorded at term-equivalent age (36–41 wk PMA) before NICU discharge. Video recordings of feeding were conducted with a lateral view of the infant's mouth and jaw. Parents and health care professionals were instructed to feed the baby how they would normally, and no instructions were given during the recording.

The 52 videos of a full oral feeding at term-equivalent age were scored using the Neonatal Eating Outcome Assessment by an evaluator trained in its use. The same videos were used to score feeding using the NOMAS assessment. NOMAS scoring was done by a different examiner, who was certified in the use of the NOMAS. Neither of the evaluators who scored the Neonatal Eating Outcome Assessment and NOMAS for concurrent validity participated as an evaluator for interrater reliability.

Statistical Analysis

For interrater reliability, an intraclass correlation (ICC) was calculated for the total Neonatal Eating Outcome Assessment scores to assess reliability across raters ([Shrout & Fleiss, 1979](#)). Fleiss's κ statistics were used to determine

the relationships of Neonatal Eating Outcome Assessment categorical scores (*normal*, *questionable*, or *feeding challenged*) across raters (Fleiss & Cohen, 1973). Fleiss's κ statistics were also calculated to determine interrater reliability of the categorical (letter) scores of each of the 19 scorable items on the Neonatal Eating Outcome Assessment. A biostatistician performed the interrater reliability analyses using R 3.2.0 (R Foundation for Statistical Computing, Vienna, Austria).

For concurrent validity, an independent-samples *t* test was used to determine associations between the NOMAS categorical scores (*disorganized* or *dysfunctional*) and the Neonatal Eating Outcome Assessment scores. The NOMAS scores were also compared with the Neonatal Eating Outcome Assessment categorical scores (*normal*, *questionable*, or *feeding challenged*) using χ^2 analysis. Analyses were performed using IBM SPSS Software (Version 23.0; IBM Corporation, Armonk, NY).

Results

Seven therapists participated in reliability training and testing. Four (57%) were occupational therapists, 2 (29%) were physical therapists, and 1 (14%) was a speech-language pathologist. The amount of experience they had ranged from 6.5 to 36 yr (mean [*M*] = 23.1, standard deviation [*SD*] = 11.4). The amount of experience in the NICU ranged from 4 to 23 yr [*M* = 12.9, *SD* = 7.9]. Five therapists (71%) were Certified Neonatal Therapists (Neonatal Therapy National Certification Board; <http://ntncb.com>). Six therapists (86%) reported that they had neonatal feeding expertise (4 occupational therapists, 1 physical therapist, and the speech-language pathologist), with the percentage of their caseload devoted to neonatal feeding ranging from 25% to 100% (*M* = 51, *SD* = 35). Two therapists (29%) had NOMAS certification, and 3 (43%) had other certifications related to feeding.

The ICC of the Neonatal Eating Outcome Assessment total continuous score was 0.90 (confidence interval [CI] [0.70, 0.99]), which is considered excellent. ICC scores of less than 0.40 are considered poor; 0.40–0.59, fair; 0.60–0.74, good; and 0.75–1.00, excellent (Cicchetti, 1994). Fleiss's κ for the total categorical score of the Neonatal Eating Outcome Assessment was 0.56 ($z = 7.79$, $p < .001$), which is considered moderate agreement. κ scores < 0 are considered to have poor agreement, 0.01–0.02 to have slight agreement, 0.21–0.40 to have fair agreement, 0.41–0.60 to have moderate agreement, 0.61–0.80 to have substantial agreement, and 0.81–1.00 to have almost perfect agreement (Landis & Koch, 1977).

The 7 therapists viewed the five feeding videos and gave letter scores for each Neonatal Eating Outcome Assessment item (Figure 2, which also includes the calculated Fleiss's κ scores). Associations between NOMAS scores and Neonatal Eating Outcome Assessment scores were observed (Table 1). Dysfunctional NOMAS scores were associated with lower Neonatal Eating Outcome Assessment continuous scores ($t[49.4] = 3.7$, *M* difference = 12.2, 95% CI [5.6, 18.8], $p = .001$). NOMAS scores were also related to Neonatal Eating Outcome Assessment categorical scores ($\chi^2[2, N = 52] = 6.40$, $p = .04$).

Discussion

The key findings of this study are that the Neonatal Eating Outcome Assessment has excellent reliability and that concurrent validity was established through the Neonatal Eating Outcome Assessment's relationship with the NOMAS. The excellent reliability of the Neonatal Eating Outcome Assessment is consistent with the previously published study assessing the reliability of Version 4 of the Neonatal Eating Outcome Assessment (Pineda et al., 2018).

Although excellent reliability on the total Neonatal Eating Outcome Assessment score was observed, some of the individual items that make up the assessment had less optimal reliability, such as swallow, quality of movements, and tone. It is possible that reliability could have been affected by lack of variability in characteristics across infants, or it could signal the need for more instruction or teaching on traits that make up each score of these items. It is also possible

Figure 2. Neonatal Eating Outcome Assessment item letter scores assigned by the therapist raters for each feeding video and associated Fleiss's κ statistics.

Item I-1: Arousal/State Organization: Physiological								Item I-2: Stability							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	D	D	D	A	D	D	D	Baby 1	C	C	D	C	D	D	D
Baby 2	D	D	C	C	D	D	D	Baby 2	D	D	D	D	D	D	D
Baby 3	D	D	D	D	D	D	D	Baby 3	D	D	D	D	D	D	D
Baby 4	D	D	D	D	D	D	D	Baby 4	D	D	D	D	D	D	D
Baby 5	D	D	D	D	D	D	D	Baby 5	D	D	D	D	D	D	D
$\kappa = 0.05$ (slight agreement), $z = 0.605$, $p = .545$								$\kappa = 0.27$ (fair agreement), $z = 2.78$, $p = .005$							
Item I-3: Respiratory Support*								Item I-4: Rooting and Grasp							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	D	D	D	D	D	D	D	Baby 1	D	D	D	B	D	D	D
Baby 2	D	D	D	D	D	D	D	Baby 2	D	D	D	D	D	D	D
Baby 3	D	D	D	D	D	D	D	Baby 3	E	C	C	D	C	C	B
Baby 4	D	D	D	D	D	D	D	Baby 4	D	D	D	D	D	D	D
Baby 5	D	D	D	D	D	D	D	Baby 5	D	D	C	B	C	D	D
$\kappa = 0.26$ (fair agreement), $z = 3.32$, $p < .001$								$\kappa = 0.26$ (fair agreement), $z = 3.72$, $p < .001$							
Item I-5: Initiation of Sucking								Item I-6: Tongue							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	C	D	D	C	D	D	D	Baby 1	D	D	B	D	D	D	D
Baby 2	D	D	D	D	D	D	D	Baby 2	D	D	B	D	D	D	D
Baby 3	B	C	C	C	B	C	B	Baby 3	D	B	B	B	B	B	E
Baby 4	D	D	D	C	D	D	D	Baby 4	D	D	D	D	D	D	D
Baby 5	D	D	C	C	C	D	D	Baby 5	D	D	E	C	B	D	D
$\kappa = 0.26$ (fair agreement), $z = 3.32$, $p < .001$								$\kappa = 0.24$ (fair agreement), $z = 3.12$, $p < .001$							
Item I-7: Nonnutritive Sucking*								Item II-1: Suck-Swallow-Breathe Coordination							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	D	D	D	D	D	D	D	Baby 1	C	C	B	B	C	D	D
Baby 2	D	D	D	D	D	D	D	Baby 2	C	D	C	D	C	D	D
Baby 3	D	D	D	D	D	D	D	Baby 3	D	B	B	B	C	B	C
Baby 4	D	D	D	D	D	D	D	Baby 4	B	B	B	B	C	B	C
Baby 5	D	D	D	D	D	D	D	Baby 5	A	C	B	A	B	B	C
$\kappa = 0.05$ (slight agreement), $z = 0.725$, $p = .47$								$\kappa = 0.05$ (slight agreement), $z = 0.725$, $p = .47$							
Item II-2: Sucking Burst Length								Item II-3: Suction							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	C	C	B	B	C	C	D	Baby 1	C	D	B	B	C	C	C
Baby 2	C	C	C	C	B	C	C	Baby 2	C	C	B	C	C	C	C
Baby 3	B	D	B	B	D	C	C	Baby 3	C	B	B	B	C	B	B
Baby 4	D	D	D	C	D	D	D	Baby 4	C	D	D	C	C	C	C
Baby 5	B	B	B	A	A	B	C	Baby 5	B	B	D	A	B	B	C
$\kappa = 0.23$ (fair agreement), $z = 3.58$, $p < .001$								$\kappa = 0.13$ (slight agreement), $z = 1.85$, $p = .06$							
Item II-4: Oral Tone								Item II-5: Quality of Sucking Movements							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	C	C	B	B	C	C	C	Baby 1	B	B	B	B	C	C	C
Baby 2	C	B	B	C	C	C	C	Baby 2	B	B	C	C	C	C	C
Baby 3	C	B	B	B	C	C	B	Baby 3	C	B	C	C	C	C	B
Baby 4	C	C	D	C	C	C	C	Baby 4	B	D	C	C	B	C	C
Baby 5	C	C	D	A	B	C	C	Baby 5	B	D	A	B	A	B	C
$\kappa = -0.01$ (poor agreement), $z = -0.18$, $p = .86$								$\kappa = 9.7e^{-17}$ (poor agreement), $z = 1.2e^{-15}$, $p = 1.0$							

Note. Asterisk indicates that statistics could not be calculated because of consistency of scoring across all infants.

that these aspects of feeding are challenging to visualize on video; therefore, it is unclear whether this challenge could have affected scores across the raters.

Therapists expressed concern after scoring the videos because of how feeding continued for some infants who were uncomfortable with the feed or who were showing stress signs. Some therapists also expressed frustration with the inability to see the infant's vital signs on the videos or see the infant's full body at different angles. In addition, there are

Figure 2. Neonatal Eating Outcome Assessment item letter scores assigned by the therapist raters for each feeding video and associated Fleiss's κ statistics. (Cont.)

Item II-6: Behavioral Response to Feeding								Item II-7: Fluid Loss							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	C	C	A	B	C	C	B	Baby 1	D	D	C	D	D	D	D
Baby 2	C	D	C	C	C	D	D	Baby 2	C	B	C	D	C	B	C
Baby 3	C	D	D	C	C	D	C	Baby 3	C	C	C	C	C	C	D
Baby 4	C	D	D	C	C	D	D	Baby 4	C	B	C	C	B	C	C
Baby 5	B	A	C	B	A	B	A	Baby 5	B	B	B	B	B	B	C
$\kappa = 0.08$ (slight agreement), $z = 1.36$, $p = .17$								$\kappa = 0.35$ (fair agreement), $z = 5.22$, $p < .001$							
Item II-8: Swallow								Item II-9: Respiratory Control During Feeding							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	D	B	B	B	B	C	C	Baby 1	B	B	B	C	B	C	B
Baby 2	C	B	C	D	B	C	C	Baby 2	C	D	C	C	B	D	C
Baby 3	C	D	C	D	C	D	C	Baby 3	D	D	D	D	C	D	D
Baby 4	C	D	C	D	B	D	D	Baby 4	C	C	D	C	D	D	D
Baby 5	C	C	B	C	B	D	B	Baby 5	C	C	C	C	B	D	B
$\kappa = -0.01$ (poor agreement), $z = -0.09$, $p = .93$								$\kappa = 0.18$ (slight agreement), $z = 2.55$, $p = .01$							
Item III-1: Feeding Completion								Item III-2: State Maintenance							
Rater:	1	2	3	4	5	6	7	Rater:	1	2	3	4	5	6	7
Baby 1	C	A	C	C	A	C	B	Baby 1	D	D	D	A	D	D	B
Baby 2	B	B	B	B	B	B	B	Baby 2	C	B	C	C	B	B	B
Baby 3	C	C	B	B	C	B	D	Baby 3	D	D	D	D	D	D	D
Baby 4	D	D	D	D	D	D	D	Baby 4	D	D	D	D	D	D	D
Baby 5	C	B	B	B	C	B	B	Baby 5	B	B	B	B	A	B	B
$\kappa = 0.45$ (moderate agreement), $z = 6.99$, $p < .001$								$\kappa = 0.47$ (moderate agreement), $z = 7.07$, $p < .001$							
Item III-3: Volume Consumed															
Rater:	1	2	3	4	5	6	7								
Baby 1	D	D	D	D	D	D	D								
Baby 2	E	E	E	E	E	E	E								
Baby 3	D	C	D	B	D	D	D								
Baby 4	F	D	F	F	F	F	F								
Baby 5	C	C	C	C	C	C	B								
$\kappa = 0.71$ (substantial agreement), $z = 13.8$, $p < .001$															

aspects of the assessment that are scored on the basis of the feeder's intervention, such as initiation of sucking or suck-swallow-breathe coordination. Some therapists found it hard to ascertain whether various cues, such as tactile stimulation, that were provided before or during the feed were necessary or purely subjective to the feeder's judgment, which may have affected the therapist's scoring. This difficulty may have made scoring especially difficult when the feeder differed in technique from what the participant would have done in the same situation.

Although reliability of each item score could be limited due to scoring from a video, which multiple participants reported as a challenge, it remains unclear how assessment reliability might change with in-person scoring. Despite these challenges, excellent reliability was achieved for the total score; moderate agreement for the categorical score, and predominately fair, moderate, and slight agreement for each item score, with some having poor agreement.

Most trainings on neonatal assessments occur in person over several days (Einspieler & Prechtl, 2005; Palmer et al., 1993; Tronick & Lester, 1996). Before reliability testing, trainings were limited to 2 hr, followed by individual practice. None of the therapists participating in the reliability testing achieved the metric of 85% agreement across all the individual items on the assessment. This finding provides insight about the potential need to expand training on use of the Neonatal Eating Outcome Assessment, which may further improve reliability.

Concurrent validity of the Neonatal Eating Outcome Assessment has not previously been reported. Associations between NOMAS scores and Neonatal Eating Outcome Assessment scores were observed, establishing concurrent validity. The NOMAS is considered to be one of the most commonly used standardized assessments of neonatal

Table 1. Associations Between NOMAS Scores and Neonatal Eating Outcome Assessment Scores

Assessment	NOMAS Disorganized (<i>n</i> = 22), <i>n</i> (%) or <i>M</i> (<i>SD</i>)	NOMAS Dysfunctional (<i>n</i> = 30), <i>n</i> (%) or <i>M</i> (<i>SD</i>)	<i>P</i> *
Neonatal Eating Outcome Assessment total score	73.9 (9.3)	61.7 (14.3)	.001
Neonatal Eating Outcome categorical score			.04
Normal (<i>n</i> = 12)	7 (58)	5 (42)	
Questionable (<i>n</i> = 26)	13 (50)	13 (50)	
Feeding challenged (<i>n</i> = 14)	2 (14)	12 (86)	

Note. There were no normal NOMAS scores in the preterm cohort; therefore, only disorganized and dysfunctional NOMAS scores are reported. *M* = mean; NOMAS = Neonatal Oral Motor Assessment Scale; *SD* = standard deviation.

*Investigation of differences between NOMAS scores and Neonatal Eating Outcome Assessment scores using independent-samples *t* test; investigation of differences between NOMAS scores and Neonatal Eating Outcome Assessment categorical scores using χ^2 analysis.

feeding, and it is the most investigated for psychometric qualities (Bickell et al., 2018). As a result of its established validity, the Neonatal Eating Outcome Assessment offers an alternative standardized feeding tool, with the additional benefit of being an assessment that can gauge normal versus abnormal feeding performance across PMA with good psychometric properties.

Study Limitations and Future Research

Limitations of this study include video evaluation of infant feeding rather than natural, face-to-face evaluation. This method required raters to make assumptions about information not visible on the video, such as swallow function. It also did not enable the

therapists to drive the feeding, including making decisions of when to start and stop the feeding process, which could have complicated scoring. All videos used for reliability testing were infants being bottle fed. Some letter scores of items, such as nonnutritive sucking and respiratory support, were difficult to interpret statistically because they lacked variability in scoring across raters and feedings. For concurrent validity, a convenience sample was used that included infants from two separate cohorts. Feeding at term-equivalent age was used to compare across the two measures, but evaluations of other time periods (PMAs) were not included.

Future research is needed on breastfed infants, with infants at different PMAs, with a sample with large variability in infant feeding performance, and with assessments conducted in person. The Neonatal Eating Outcome Assessment has promising clinical utility and interrater reliability and concurrent validity, but further modifications to training may better prepare raters to administer and score the assessment appropriately.

Implications for Occupational Therapy Practice

This study has the following implications for occupational therapy practice:

- Few tools exist for assessing oral feeding; however, having valid and reliable assessments to determine problems with early oral feeding is important and can drive targeted interventions to optimize outcomes.
- This article demonstrates psychometric qualities of the Neonatal Eating Outcome Assessment, which is important for clinical practice because clinicians can benefit from having an expanded repertoire of standardized tools to assess oral feeding in preterm infants in the NICU.
- Standardized assessments enable consistent documentation to determine the effects of interventions as well as for conduct of research to build the evidence to support practice.
- Early oral feeding assessment can aid early identification of feeding difficulties, enabling earlier therapy referrals and interventions.

Conclusion

With very few structured, standardized assessments available for neonatal feeding, this tool can improve early detection of infant feeding problems and aid in identifying the need for early services to improve long-term outcomes. The Neonatal Eating Outcome Assessment's reliability is excellent, and it has concurrent validity, because scores relate to the NOMAS. Future work is needed to determine the stability of scores across PMA and to evaluate the predictive

validity of the Neonatal Eating Outcome Assessment. Psychometric investigations of the abbreviated form, for infants not yet orally feeding, and for infants who are breastfeeding are also needed. ■

References

- Barlow, S. M. (2009). Oral and respiratory control for preterm feeding. *Current Opinion in Otolaryngology and Head and Neck Surgery*, 17, 179–186. <https://doi.org/10.1097/MOO.0b013e32832b36fe>
- Bertoncelli, N., Cuomo, G., Cattani, S., Mazzi, C., Pugliese, M., Coccolini, E., . . . Ferrari, F. (2012). Oral feeding competences of healthy preterm infants: A review. *International Journal of Pediatrics*, 2012, 896257. <https://doi.org/10.1155/2012/896257>
- Bickell, M., Barton, C., Dow, K., & Fucile, S. (2018). A systematic review of clinical and psychometric properties of infant oral motor feeding assessments. *Developmental Neurorehabilitation*, 21, 351–361. <https://doi.org/10.1080/17518423.2017.1289272>
- Braun, M., & Palmer, M. M. (1985). A pilot study of oral–motor dysfunction in “at-risk” infants. *Physical and Occupational Therapy in Pediatrics*, 5, 13–26. https://doi.org/10.1080/J006v05n04_02
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- da Costa, S. P., Hübl, N., Kaufman, N., & Bos, A. F. (2016). New scoring system improves inter-rater reliability of the Neonatal Oral–Motor Assessment Scale. *Acta Paediatrica*, 105, e339–e344. <https://doi.org/10.1111/apa.13461>
- da Costa, S. P., van den Engel-Hoek, L., & Bos, A. F. (2008). Sucking and swallowing in infants and diagnostic tools. *Journal of Perinatology*, 28, 247–257. <https://doi.org/10.1038/sj.jp.7211924>
- da Costa, S. P., & van der Schans, C. P. (2008). The reliability of the Neonatal Oral–Motor Assessment Scale. *Acta Paediatrica*, 97, 21–26. <https://doi.org/10.1111/j.1651-2227.2007.00577.x>
- de Kieviet, J. F., Piek, J. P., Aarnoudse-Moens, C. S., & Oosterlaan, J. (2009). Motor development in very preterm and very low-birth-weight children from birth to adolescence: A meta-analysis. *JAMA*, 302, 2235–2242. <https://doi.org/10.1001/jama.2009.1708>
- Dodrill, P. (2011). Feeding difficulties in preterm infants. *Infant, Child, and Adolescent Nutrition*, 3, 324–331. <https://doi.org/10.1177/1941406411421003>
- Einarsson-Backes, L. M., Deitz, J., Price, R., Glass, R., & Hays, R. (1994). The effect of oral support on sucking efficiency in preterm infants. *American Journal of Occupational Therapy*, 48, 490–498. <https://doi.org/10.5014/ajot.48.6.490>
- Einspieler, C., & Prechtl, H. F. R. (2005). Prechtl's Assessment of General Movements: A diagnostic tool for the functional assessment of the young nervous system. *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 61–67. <https://doi.org/10.1002/mrdd.20051>
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613–619. <https://doi.org/10.1177/001316447303300309>
- Gennattasio, A., Perri, E. A., Baranek, D., & Rohan, A. (2015). Oral feeding readiness assessment in premature infants. *American Journal of Maternal/Child Nursing*, 40, 96–104. <https://doi.org/10.1097/nmc.0000000000000115>
- Gewolb, I. H., Vice, F. L., Schwietzer-Kenney, E. L., Taciak, V. L., & Bosma, J. F. (2001). Developmental patterns of rhythmic suck and swallow in preterm infants. *Developmental Medicine and Child Neurology*, 43, 22–27. <https://doi.org/10.1017/S0012162201000044>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42, 377–381.
- Howe, T. H., Sheu, C. F., Hsieh, Y. W., & Hsieh, C. L. (2007). Psychometric characteristics of the Neonatal Oral–Motor Assessment Scale in healthy preterm infants. *Developmental Medicine and Child Neurology*, 49, 915–919. <https://doi.org/10.1111/j.1469-8749.2007.00915.x>
- Jensen, D., Wallace, S., & Kelsay, P. (1994). A new breastfeeding assessment tool. *Journal of Human Lactation*, 10, 9–10. <https://doi.org/10.1177/089033449401000119>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>
- Ludwig, S. M., & Waitzman, K. A. (2007). Changing feeding documentation to reflect infant-driven feeding practice. *Newborn and Infant Nursing Reviews*, 7, 155–160. <https://doi.org/10.1053/j.nainr.2007.06.007>
- Matthews, M. K. (1988). Developing an instrument to assess infant breastfeeding behaviour in the early neonatal period. *Midwifery*, 4, 154–165. [https://doi.org/10.1016/S0266-6138\(88\)80071-8](https://doi.org/10.1016/S0266-6138(88)80071-8)
- McGrath, J. M., & Braescu, A. V. (2004). State of the science: Feeding readiness in the preterm infant. *Journal of Perinatal and Neonatal Nursing*, 18, 353–368. <https://doi.org/10.1097/00005237-200410000-00006>
- Mulford, C. (1992). The Mother–Baby Assessment (MBA): An “Apgar score” for breastfeeding. *Journal of Human Lactation*, 8, 79–82. <https://doi.org/10.1177/089033449200800216>
- Nyqvist, K. H., Rubertsson, C., Ewald, U., & Sjöden, P. O. (1996). Development of the Preterm Infant Breastfeeding Behavior Scale (PIBBS): A study of nurse–mother agreement. *Journal of Human Lactation*, 12, 207–219. <https://doi.org/10.1177/089033449601200318>
- Palmer, M. M., Crawley, K., & Blanco, I. A. (1993). Neonatal Oral–Motor Assessment Scale: A reliability study. *Journal of Perinatology*, 13, 28–35.
- Palmer, M. M., & Heyman, M. B. (1999). Developmental outcome for neonates with dysfunctional and disorganized sucking patterns: Preliminary findings. *Infant–Toddler Intervention: The Transdisciplinary Journal*, 9, 299–308.

- Pineda, R., Dewey, K., Jacobsen, A., & Smith, J. (2019). Non-nutritive sucking in the preterm infant. *American Journal of Perinatology*, 36, 268–276. <https://doi.org/10.1055/s-0038-1667289>
- Pineda, R., Harris, R., Foci, F., Roussin, J., & Wallendorf, M. (2018). Neonatal Eating Outcome Assessment: Tool development and inter-rater reliability. *Acta Paediatrica*, 107, 414–424. <https://doi.org/10.1111/apa.14128>
- Ross, E. S., & Browne, J. V. (2002). Developmental progression of feeding skills: An approach to supporting feeding in preterm infants. *Seminars in Neonatology*, 7, 469–475. <https://doi.org/10.1053/siny.2002.0152>
- Shrago, L., & Bocar, D. (1990). The infant's contribution to breastfeeding. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 19, 209–215. <https://doi.org/10.1111/j.1552-6909.1990.tb01638.x>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Thoyre, S. M., Shaker, C. S., & Pridham, K. F. (2005). The Early Feeding Skills Assessment for preterm infants. *Neonatal Network*, 24, 7–16. <https://doi.org/10.1891/0730-0832.24.3.7>
- Tronick, E. Z., & Lester, B. M. (1996). The NICU Network Neurobehavioral Scale: A comprehensive instrument to assess substance-exposed and high-risk infants. *NIDA Research Monograph*, 166, 198–204.
- Zarem, C., Kidokoro, H., Neil, J., Wallendorf, M., Inder, T., & Pineda, R. (2013). Psychometrics of the Neonatal Oral Motor Assessment Scale. *Developmental Medicine and Child Neurology*, 55, 1115–1120. <https://doi.org/10.1111/dmcn.12202>

Roberta Pineda, PhD, OTR/L, is Assistant Professor, Program in Occupational Therapy and Department of Pediatrics, Washington University School of Medicine, St. Louis, MO; pineda_r@kids.wustl.edu

Lara Liszka, OTD/S, is Doctor of Occupational Therapy Student, Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO.

Jenny Kwon, OTD/S, is Doctor of Occupational Therapy Student, Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO.

Michael Wallendorf, PhD, is Research Statistician, Division of Biostatistics, Washington University School of Medicine, St. Louis, MO.

Acknowledgments

We thank Katie Ross, Madison Rolling, Jessica Roussin, Maggie Crabtree, Molly Grabill, Danielle Prince, Margaux Collins, Pido Tran, Kylie Buma, Sarah TenBarge, and Tiffany Le. We also thank the therapists who participated in the reliability testing as well as the infants and families whose participation made this research possible. This work was supported by National Institutes of Health R24 (5R24HD06568805) awarded by the Boston Rehabilitation Outcomes Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U54 HD087011) to the Intellectual and Developmental Disabilities Research Center at Washington University.