

Reliability and Validity Testing of Three Breastfeeding Assessment Tools

Janice M. Riordan, ARNP, EdD, FAAN, Mary Koehn, RN, MSN, FACCE

Objective: This study examined validity and reliability of three clinical instruments that assess feedings at the breast.

Design: A descriptive correlational design testing the validity and interrater and test-retest reliability of instruments.

Setting: Hospital rooms and the participants' homes.

Subjects: Eleven breastfeeding women and their neonates were videotaped in 23 breastfeeding observations.

Interventions: The Infant Breastfeeding Assessment Tool (IBFAT), the Mother Baby Assessment Tool (MBA), and the LATCH assessment tool were scored by three nurse raters using videotapes of breastfeedings. Instruments were completed twice by each rater with a 6-month period between administration.

Main Outcome Measures: To test validity, test-retest, and interrater reliability, Spearman correlation coefficients among raters' breastfeeding assessment scores, among scores of each instrument, and between test and retest scores of raters. Percent of agreement among raters for each of the items in the three tools.

Results: Reliability coefficients for all three assessment tools are below acceptable levels for clinical decisions. Spearman rank coefficients of pairwise interrater correlations were .57, .27, and .69 for the IBFAT; .66, .64, and .33 for the MBA; and .11, .46, and .48 for the LATCH assessment tool. Spearman rank coefficients among instrument scores were .69, .78, and .68. Test-retest correlations were .88, .78, and .64. Percent of agreement among raters for each of the items in the three tools was highly variable, ranging from 37.0 to 97.2.

Conclusion: The IBFAT, MBA, and LATCH as tools to measure breastfeeding effectiveness are not

sufficiently reliable at this stage in their development; thus, these tools cannot be valid for clinical use. These tools need to be revised and retested before use in clinical practice to identify breastfeeding mother-infant pairs who need intervention. *JOGNN*, 26, 181-187; 1997.

Accepted: April 1996

Early hospital discharge of new mothers presents challenges for postpartum nurses, including assessment of breastfeeding skills. Shorter hospital stays limit periods of observation and time for patient teaching. New mothers return home with less experience and information than was garnered previously during longer postpartum stays. The availability of a simple, reliable tool to assess breastfeedings and therefore predict breastfeeding problems that need follow-up would be a valuable contribution to the care of mothers and their newborns.

The availability of a simple, reliable tool to assess breastfeedings and predict breastfeeding problems that need follow-up would be a valuable contribution to the care of mothers and their newborns.

In addition to identifying mother-infant pairs at risk for breastfeeding problems, breastfeeding assessment tools can serve as indicators for

quality assurance measures and as research variables. Three tools that systematically assess and score feedings at the breast, the Infant Breastfeeding Assessment Tool (Matthews, 1988), the Mother Baby Assessment Tool (Mulford, 1992), and the LATCH assessment tool (Jensen, Wallace, & Kelsay, 1994) have been developed independently by maternity nurses and are described in detail below. These assessment tools attempt to quantify complex infant and maternal behaviors during breastfeeding, and their numerical scores are intended to indicate the efficacy of the feeding observed. The purpose of this study was to describe reliability of these tools and the extent to which they measure breastfeeding effectiveness so they can be accepted for clinical use.

Reliability is concerned with how consistently the measurement technique or tool measures the concept of interest or a specific behavior. Reliability testing should answer the following questions. When two or more nurses assess a feeding using one of the measures, are they likely to have the same or similar scores (interrater reliability)? Would the results be the same if the assessment measure were repeated (test-retest)? Evidence of interrater reliability is based on documentation of high positive correlations between all pairs of raters who are observing a target behavior, in this case breastfeeding (Nunnally, 1978).

Another measure of objectivity is the validity of the tool. Validity is concerned with determining the extent to which the measurement technique of the tool actually reflects the concept of interest or a specific behavior. One approach to evaluating for validity involves showing that different tools that measure the same behavior converge or are positively correlated (Nunnally, 1978). If these three tools are presumed to measure the same construct (effective breastfeeding) as referent or exemplars of the construct, a high positive correlation between the scores on each is predicted and is called construct validity (Cronbach & Meehl, 1955).

This study, which focused on assessment of the Infant Breastfeeding Assessment Tool (IBFAT), the Mother Baby Assessment Tool (MBA), and the LATCH assessment tool for interrater reliability and construct validity, provides initial data needed to determine the adequacy of these tools for use in clinical practice. When these tools have been demonstrated to possess these properties, it will be possible to ascertain whether any or all are accurate predictors of adequate infant intake, especially during the first few days or weeks after hospital discharge.

Background

Neonatal suckling behavior is a measure of behavioral organization. Effective suckling behavior suggests that an infant has achieved the maturity required for ef-

fective oral feedings and subsequent survival (Medoff-Cooper & Ray, 1995). Rooting and suckling are influenced by infant age and state of wakefulness (Anderson, McBride, Dahm, Ellis, & Vidyasager, 1982; Gentry & Aldrich, 1948; Pollitt, Consolazio, & Goodkin, 1981; Weber, Woolridge, & Baum, 1986). The neonate's ability to suckle and to root progressively matures after birth (Bowen-Jones, Thompson, & Drewitt, 1982; Drewitt & Woolridge, 1979; Selley, Ellis, & Flack, 1986) and becomes more organized (Medoff-Cooper, 1991). When the infant is feeding at the breast, the tongue is grooved around and under the nipple and moves in a peristaltic action from front to back (Marmet & Shell, 1984; Woolridge, 1986). Both nutritive and nonnutritive suckling occur throughout the feeding, and they may not appear to be distinctly different.

Objective measurement of feeding at the breast is difficult because complex interactive behaviors are required of two individuals. Generally, rooting, active suckling, signs of swallowing, maternal milk ejection, signs of satiety, adequate infant "positioning" (grasps sufficient areolar tissue), and maternal verbalization of satisfaction with breastfeeding have served as global components of effective breastfeedings (Gentry & Aldridge, 1948; Kron, Stein, & Goddard, 1966; Lethbridge, McClurg, Henrikson, & Wall, 1993; Matthews, 1988; Riordan, 1993; Weber et al., 1986; Woolridge, 1986). Nipple structure also may affect suckling. Retraction or inversion can prevent the infant from effectively milking the lactiferous sinuses that lie beneath the areola (Jensen, Wallace, & Kelsey, 1994; Riordan, 1993).

Breastfeeding assessment tools are at a beginning stage in their development, and health professionals do not agree on what components comprise an effective breastfeeding.

Breastfeeding assessment tools are at a beginning stage in their development, and health professionals do not yet agree about what components comprise effective breastfeeding. The accuracy of observing these components and then estimating volume intake had not been tested until Meier and colleagues (1994) found that using clinical clues to estimate how much breastmilk the infant ingested was not an accurate indicator of intake for pre-term infants.

TABLE 1
Comparison of Breastfeeding Assessment Tools

<i>Characteristic</i>	<i>IBFAT</i>	<i>MBA</i>	<i>LATCH</i>
Focus on	Infant	Infant and mother	Infant and mother
Scored by	Mother or nurse	Nurse	Mother or nurse
Time frame	Progressive: beginning to ending	Progressive: beginning to ending	Static
Analysis of sequential scores	Use mean of scores	Use best of scores	Expect increase in scores
Measures	Signaling Rooting Suckling	Readiness Position Latch-on; milk transfer Outcome	Latch-on Audible swallowing Nipple comfort Assistance needed with positioning

IBFAT = Infant Breastfeeding Assessment Tool; MBA = Mother Baby Assessment Tool; LATCH = breastfeeding assessment tool in which "L" denotes how well the infant latches onto the breast, "A" denotes the amount of audible swallowing noted, "T" denotes the mother's nipple type, "C" denotes the mother's level of comfort, and "H" denotes the amount of help the mother needs to hold her infant to the breast.

Method

Sample

Twenty-eight feedings at the breast were videotaped using a convenience sample of 13 breastfeeding mothers and their neonates in a Midwestern city hospital. The study was explained, and informed consent was obtained. The mothers were white and ranged from ages 20 to 36 years. All infants were healthy, full-term, and had Apgar scores at 5 minutes of 9 or above. Participants were experienced and inexperienced mothers with different breastfeeding experiences. Eight had breastfed previously.

Instruments

The IBFAT is a short assessment and measurement of infant breastfeeding competence (Matthews, 1988, 1991). The IBFAT represents four major components of infant breastfeeding behavior: (a) readiness to feed, (b) rooting, (c) fixing, and (d) suckling. The range of scores for each of the four components is 0–3. A total score can range from 0 to 12, with 12 being the score for vigorous, effective suckling. The instrument was based on observations in clinical practice, knowledge obtained from a literature review, and consultation with experts in nursing and child development. The IBFAT also measures the mother's perception of, and satisfaction with, the feeding. Matthews (1988) found that interrater reliability between mothers' and researchers' assessments was 91%. More recent research indicated that mothers who reported low breastfeeding satisfaction were mothers of infants who were rated low on the IBFAT (Goer, Humenick, & Hill, 1994).

LATCH is a breastfeeding charting system that provides a systematic method for gathering information about individual breastfeeding sessions (Jensen et al., 1994). The *LATCH* assessment tool was developed to identify areas of needed intervention and to determine priorities in providing patient care and teaching. Because not all breastfeedings can be observed by the hospital staff, mother-reported scores are recorded for continuing assessment. The system assigns a numerical score, (0, 1, or 2) to five key components of breastfeeding for a possible total score of 10. Each letter of the acronym *LATCH* denotes an area of assessment. "L" is for how well the infant latches onto the breast. "A" is for the amount of audible swallowing noted. "T" is for the mother's nipple type. "C" is for the mother's level of comfort. "H" is for the amount of help the mother needs to hold her infant to the breast.

The Mother-Baby Assessment Tool is meant to assess the process of learning to breastfeed. It divides the process of breastfeeding into five steps: (a) signaling, (b) positioning, (c) fixing, (d) milk transfer, and (e) ending (Mulford, 1992). For each step, there is both a mother and an infant behavior, based on the assumption that breastfeeding is a mutual effort. Ten is the highest possible score: 5 for maternal behaviors and 5 for infant behaviors in each of the steps indicates a highly effective feeding. No evidence of reliability or validity of the MBA has been published.

Table 1 compares the similarities and differences among these three tools. Each tool is different in its use of components to represent the construct, "effective breastfeeding." Audible swallowing is measured in the MBA and the *LATCH* assessment tool, but it is not rep-

resented in the IBFAT. Positioning is used in the MBA and the LATCH assessment tool, an indication that the infant is adequately grasping the nipple and areola, whereas it is not in the IBFAT. Only the LATCH assessment tool addresses the mother's nipple as an indicator of effective breastfeeding.

Procedure

Videotapes were taken of 28 breastfeedings using 13 mother-infant dyads. Because first feedings usually are not as effective as later feedings, the feedings were taped at various times during the 1st week postpartum. The investigators thought that conducting the videotaping sessions over a week would illustrate a range of feeding effectiveness as infants became progressively more efficient at feeding. The timing of the taping was random, depending on the convenience of the mother and the nurse who was videotaping the feedings. The number of individual feedings taped for each mother ranged from 1 to 3; the average was 2.

The videotape machine was mounted on a tripod at a level above the mother and infant approximately two feet away from the mother's head. The same equipment, a video recorder (Sony CCD-F40, 1988), was used for all subjects. A media specialist edited out unusable sections (40 minutes of 200 minutes) where breastfeeding could not be observed or where the mother's face was evident.

The completed recordings, each segment approximately 7 minutes long, were compiled onto videotapes. Each taped feeding then was rated independently by three maternity nurses skilled in breastfeeding. Before scoring began, raters were oriented to the three breastfeeding assessment tools and the scoring protocol and given the same instructions about their use. No other specific training was provided. The scoring system (ordinal rating scales) was given to all raters on a printed sheet.

Raters did not view any videotapes before beginning the rating procedure. The segments were shown according to the order of when the videotapes were made: The first taping was shown first and the last taping shown last. Without discussion, the nurses scored each observation simultaneously and independently on the rating forms. Raters were asked to record impressions about the tools in addition to their scored breastfeeding observations.

To examine test-retest reliability, 6 months after the original data were collected, a random sample of 12 taped feeding observations was replayed for rating by the original three nurse raters, who were blinded to their original assessments.

Individual taped observations rather than participants were used as units of analysis. Pearson correlation coefficients were used to relate observation scores among

the three breastfeeding assessment tools because the tool scores ranged from 0 to 10. Spearman correlation coefficients then were calculated to determine the correlations between rater scores. Agreement among raters was recorded as the percentage of the ratings on each tool that are in agreement across the three raters. Because potential harm can result from unreliable breastfeeding assessment scores, the level of 80% was designated as the minimum level of agreement that could be tolerated without jeopardizing patient safety.

Results

Five of the 28 observations were of insufficient quality for inclusion in the study. Thus, 23 observations were rated by three raters using three instruments. Estimates of interrater reliability were obtained using Spearman correlation coefficients calculated for all pairwise ratings. These estimates ranged from .27 to .69 for the IBFAT, .11 to .46 for the LATCH assessment tool, and .33 to .66 for the MBA (See Table 2). Pairwise, Spearman correlation coefficients to related scores among the three breastfeeding tools ranged from .68 to .78 (see Table 3). Spearman correlations between raters' scores of the first scoring and retested scoring on 12 randomly selected feedings 6 months later were .88, .78, and .64. For the retest, percentages of agreement among raters for each item of the tools were highly variable, ranging from

TABLE 2
Spearman Rank-Order Correlation Coefficients Among Raters of Breastfeedings Using Three Breastfeeding Assessment Tools (n = 23)

	Rater 2	Rater 3
IBFAT		
Rater 1	.57	.27
Rater 2		.69
LATCH		
Rater 1	.11	.46
Rater 2		.28
MBA		
Rater 1	.66	.64
Rater 2		.33

Note. IBFAT = Infant Breastfeeding Assessment Tool; MBA = Mother Baby Assessment Tool; LATCH = breastfeeding assessment tool in which "L" denotes how well the infant latches onto the breast, "A" denotes the amount of audible swallowing noted, "T" denotes the mother's nipple type, "C" denotes the mother's level of comfort, and "H" denotes the amount of help the mother needs to hold her infant to the breast.

TABLE 3
Pairwise Spearman Rank-Order Correlations of Instrument Scores for 23 Observations of 3 Tools by 3 Raters (N = 69)

<i>Instrument</i>	<i>r</i>
LATCH and IBFAT	.69
MBA and IBFAT	.78
LATCH and MBA	.68

Note. IBFAT = Infant Breastfeeding Assessment Tool; MBA = Mother Baby Assessment Tool; LATCH = breastfeeding assessment tool in which "L" denotes how well the infant latches onto the breast, "A" denotes the amount of audible swallowing noted, "T" denotes the mother's nipple type, "C" denotes the mother's level of comfort, and "H" denotes the amount of help the mother needs to hold her infant to the breast.

37.0% to 97.2% (see Table 4). Because there were three levels for each behavior, there was a 33.3% chance that the raters would select the same answer by chance alone. Average rater agreement for each tool were MBA, 84%; LATCH, 78%; and IBFAT, 77%.

Because 80% was the cutoff point for an acceptable minimum agreement, "milk transfer" and "outcome" in the MBA were below this minimum. Three behaviors in the IBFAT, "infant wakefulness," "need for stimulation," and "suckling effectiveness," were below 80%, whereas only one behavior in the LATCH assessment tool, "assistance needed in holding the infant," fell below that percent.

Discussion and Implications

Perhaps the most striking finding was a lack of consistency in the way the raters rank-ordered the ratings. It was expected that because they were measuring the exact same feeding at the breast on videotape, score correlations would be higher. These results indicate that the IBFAT, LATCH assessment tool, and MBA are not sufficiently reliable in their current form for clinical practice. Although there is no magical cutoff point that prevents a tool from being considered reliable, a correlation of .80 is considered a minimum value. When clinical decisions are made based on a tool, its reliability and validity should be very high, at least .90 (Nunnally, 1978). However, there is no hard and fast rule for interrater reliability of a tool. Although disappointing, this finding

TABLE 4
Percent of Agreement Among Raters of Breastfeeding Behaviors (n = 12)

<i>Tools</i>	<i>Percent Agreement</i>
IBFAT	
Wakefulness of infant before feeding	76.5
Did the infant need to be stimulated before being fed?	62.0
Presence of rooting	84.7
Length of time between placing infant at breast and start of suckle	87.5
Suckling effectiveness	77.2
LATCH	
Suckling effectiveness	84.2
Audible swallowing	95.7
Type of nipple	97.2
Comfort (breast/nipple)	88.2
Assistance needed in holding infant	54.2
MBA	
Readiness to feed (infant and mother)	95.2
Positioning (alignment, rooting, opens mouth wide)	92.5
Latch-on (latching and suckling)	89.7
Milk transfer (audible swallows, dripping milk, mother feels thirst, cramps, increased lochia, breast tingling, sleepiness)	37.0
Outcome (infant releases spontaneously, no nipple soreness)	77.0

is not surprising considering that these instruments were developed recently and their authors have not had the opportunity to revise them based on recommendations from reliability and validity testing.

Perhaps the most striking finding was a lack of consistency in the way the ratings were rank-ordered by the raters. It was expected that because they were measuring the exact same feeding at the breast, score correlations would be higher.

Matthews' (1988) reported 91% agreement between IBFAT nurse ratings and mother ratings is much higher than that found in this study. Differences in data collection may be the reason. Scores may be affected when a mother rates her own feedings while in close contact with a nurse who is also rating the same breastfeedings, whereas the videotape avoids any contact with the mother-infant pair. Conversely, direct observation data may be more precise. For example, a nurse observing videotapes cannot fully assess a mother's nipples nor ask if she felt evidence of milk transfer (thirst, cramps, and increased lochia).

When asked to critique the content of the tools, raters' comments contrasted with the quantitative results. For example, two of the raters were concerned that the LATCH assessment tool gave a high score for a soft and nontender breast compared with a lower score for a filling breast. They thought that during the 1st week postpartum, a full breast that is somewhat tender would be a positive sign that things were going well. Even so, the percentage of agreement for this item was relatively high—88.2%.

Conversely, raters noted that MBA milk transfer indicators (thirst, increased lochia, breast tingling, and sleepiness) are difficult to evaluate in a clinical setting where the nurse has only a few minutes to assess the breastfeeding. Raters agreed barely above chance, 37% of the time, on this item.

None of these tools appears to be sufficiently reliable as now written. Because a prerequisite to validity is reliability, an unreliable tool cannot be a valid measure of effective breastfeeding. Thus, several approaches can be taken to further develop and test these tools. First, this study should be replicated in a clinical setting where the mother and nurse both observe the feeding "in per-

son" and their scores are compared. Second, the authors of these tools can reevaluate and revise them before further reliability and validity testing is undertaken. A third approach is to test validity of these tools by comparing assessment tool scores using test weights as a means of comparing constructs of "breastfeeding effectiveness" and "breastmilk intake."

In the current early discharge environment, assessing infant feedings at the breast is essential for identifying early feeding problems and for consequent intervention. In addition, valid and reliable feeding assessment instruments can serve as quality assurance measures of nursing care. As the dependent variable for studies, breastfeeding assessment tools potentially can provide answers to many research questions. Without such tools, many questions about lactation cannot be posed nor their answers found.

Acknowledgment

This research was supported by a grant from Wichita State University, School of Nursing, Wichita, KS.

REFERENCES

- Anderson, G. C., McBride, M. R., Dahm, J., Ellis, M. K., & Vidyasager, D. (1982). Development of suckling in term infants from birth to four hours postbirth. *Research in Nursing and Health*, 5, 21–27.
- Bowen-Jones, A., Thompson, C., & Drewitt, R. F. (1982). Milk flow and sucking rates during breastfeeding. *Developmental Medicine and Child Neurology*, 24, 626–633.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Drewitt, R. F., & Woolridge, M. (1979). Sucking patterns of human babies on the breast. *Early Human Development*, 34, 315–320.
- Gentry, E. F., & Aldrich, C. A. (1948). Rooting reflex in the newborn infant: Incidence and effect on it of sleep. *American Journal of Diseases of Children*, 75, 528–539.
- Goer, M. W., Humenick, S., & Hill, P. (1994). Characterizations and psychoneuroimmunologic implications of secretory immunoglobulin A and cortisol in preterm and term breast milk. *Journal of Perinatal and Neonatal Nursing*, 7(4), 42–51.
- Jensen, D., Wallace, S., & Kelsay, P. (1994). LATCH: A breastfeeding charting system and documentation tool. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 23, 27–32.
- Kron, R. E., Stein, M., & Goddard, K. E. (1966). Newborn sucking behaviour affected by obstetric medication. *Pediatrics*, 37, 1012–1016.
- Lethbridge, D. J., McClurg, V., Henrikson, M., & Wall, G. (1993). Validation of the nursing diagnosis of ineffective

- breastfeeding. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 22, 57-63.
- Marmet, C., & Shell, E. (1984). Training neonates to suck correctly. *MCN: American Journal of Maternal Child Nursing*, 9, 401-407.
- Matthews, M. K. (1988). Developing an instrument to assess infant breastfeeding behaviour in the early neonatal period. *Midwifery*, 4, 154-65.
- Matthews, M. K. (1991). Mothers' satisfaction with their neonates' breastfeeding behaviors. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 20, 49-55.
- Medoff-Cooper, B. (1991). Changes in nutritive sucking patterns with increasing gestational age. *Nursing Research*, 40, 245-247.
- Medoff-Cooper, B., & Ray, W. (1995). Neonatal sucking behavior. *Image: Journal of Nursing Scholarship*, 27, 195-199.
- Meier, P., Engstrom, J. L., Crichton, C. L., Clark, D. R., Williams, M. M., & Mangurten, H. H. (1994). A new scale for in-home test-weighing for mothers of preterm and high risk infants. *Journal of Human Lactation*, 10(3), 163-168.
- Mulford, C. (1992). The mother-baby assessment (MBA): An "Apgar Score" for breastfeeding. *Journal of Human Lactation*, 8(2), 79-82.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pollitt, E., Consolazio, B., & Goodkin, G. (1981). Changes in nutritive sucking during a feed in two-day and three-day old infants. *Early Human Development* 5, 201-210.
- Riordan, J. (1993). Anatomy and psychophysiology of lactation. In J. Riordan & K. Auerbach, *Breastfeeding and human lactation* (pp. 92-93). Boston: Jones & Bartlett.
- Selley, W. G., Ellis, R. E., & Flack, F. C. (1986). Ultrasonographic study of sucking and swallowing by newborn infants. *Developmental Medicine and Child Neurology*, 28, 821-823.
- Weber, F., Woolridge, M. W., & Baum, J. D. (1986). An ultrasonographic study of the organization of suckling and swallowing by newborn infants. *Developmental Medicine and Child Neurology* 28, 19-24.
- Woolridge, M. (1986). The 'anatomy' of infant sucking. *Midwifery*, 2, 164-171.

Janice Riordan is an associate professor of nursing, Wichita State University, School of Nursing, Wichita, KS.

Mary Koehn is a clinical educator, Wichita State University, School of Nursing, Wichita KS.

Address for correspondence: Janice Riordan, EdD, FAAN, School of Nursing, Wichita State University, 1845 N. Fairmount, Box 41, Wichita, KS 67260-0041; e-mail address: jriordan@feist.com.