

Critical Review: Exploring the inter-rater reliability of two assessment tools used to identify feeding problems in neonates

Gillian Lucas

M.Cl.Sc (SLP) Candidate

University of Western Ontario: School of Communication Sciences and Disorders

This critical review examines the inter-rater reliability of two

assessment tools used in the neonatal population.
The secondary objective is to propose evidence-

while kappa values ranged from moderate to almost perfect (0.47 to 1.0). Percentage of agreement among observer B and mothers ranged from 0.60 to 0.90, whereas kappa values varied from fair to substantial (0.29 to 0.74). However, as a result of the lack of methodological rigor in this study, this research presents suggestive evidence of the inter-rater reliability of the PIBBS.

A follow-up study by Nyqvist, Sjoden, and Ewald (1999) strived to further describe the development of preterm infants' behaviour during breastfeeding, until full oral feeding and discharge home was achieved. A prospective, descriptive design was used to study 71 mother-infants pairs, while mothers functioned as data collectors using the PIBBS during observations. A scoring system was added to the scale for purpose of the study. Inter-rater agreement was examined through joint observations of the main author (main observer) and mothers, as well as between the main author and another nurse/research assistant. Seventy simultaneous assessments, one to four per mother-infant pair (n=41), were performed by observer A and mothers. These resulted in excellent agreement for nominal scale items (e.g. rooting, areolar grasp), with kappa values ranging from 0.77-0.94. Twenty-nine simultaneous observations, one to four per mother-infant pair (n=21) were made between observers A and B. These yielded good or excellent agreement for nominal scale items, and kappa values of 0.68-0.84. Narrow confidence intervals and moderate standard deviations were reported for parametric data. Yet, due to the absence of methodological rigor in this study, this research demonstrates suggestive evidence of the inter-rater reliability of the PIBBS.

A systematic review of the psychometric properties of feeding assessment tools used in neonates was carried out by Howe et al. in 2008. The purpose of their study was to comprehensively review and compare the psychometric properties of current clinical feeding assessment tools used in the neonatal population. In total, 941 articles were reviewed. The results indicate that none of the psychometric properties of the seven assessment tool groups identified were satisfactory, and limited representativeness of the samples of the research was observed in all tools. Overall, this study displays suggestive evidence of the inter-rater reliability of both the NOMAS and PIBBS.

Discussion

Appraisal of the Results

Since the body of literature investigating the inter-rater reliability for both the NOMAS and PIBBS is limited, the veracity of the scientific soundness of both instruments is called into question. Many methodological issues need to be taken into consideration when evaluating the evidence.

Participants Selection

Since

reliability measures will be more accurate as the total variance increases (Portney & Watkins, 2000, p. 559, as cited in Howe et al., 2008). As a result, it is imperative that studies which examine the reliability of an assessment measure include patients that vary in degree of functioning, from normal all the way to severely impaired. None of the NOMAS or PIBBS studies included patients of adequate variability in levels of functioning that can be said to be fully representative of the range from normal to severe.

The NOMAS investigation completed by Palmer and colleagues (1993) excluded infants with structural defects from the sample, and infants who

criteria (e.g. nonnutritive sucking was not observed) were not included for evaluation. Despite the fact that da Costa and van der Schans (2008) attempted to include a more variable sample of participants (i.e., inclusion of a control group of 23 full-term neonates), infants with particular medical conditions were still excluded from the sample (e.g. infants with multiple congenital disorders, among others). The excluded patients in both NOMAS studies likely represented a population who may potentially be at most risk for feeding difficulties, and whom would benefit greatly from early

meeting inclusion criteria, the two PIBBS studies did not. Instead, investigations by Nyqvist and collaborators (1996; 1999) included a sampling procedure that was based solely on the convenience of the researcher (i.e., a convenience sample). There was no mention in either of the PIBBS articles as to whether or not the researchers made any attempt to make certain that the sample was an accurate representation of the desired population. Instead, the researchers included individuals on the basis of availability. Due to the fact that an unknown portion of the population was excluded, bias is likely present in the convenience sample, and the degree to which the sample is actually representative of the entire population of neonates cannot be discerned (Lunsford & Lunsford, 1995).

Small sample sizes were also noted in three out of the four reviewed studies; $n=35$ for Palmer et al. (1993) NOMAS study, $n=24$, $n=10$ and $n=41$, $n=21$ 1996 and 1999 studies, respectively. Smaller samples, (i.e., $n<30$), are less likely to be acceptable representations of population characteristics (Howe et al., 2008), and power is said to be significantly reduced (Portney & Watkins, 2000, as cited in Howe et al, 2008). The NOMAS investigation completed by da Costa & van der Schans (2008) was the only study that included a relatively larger sample size ($n=75$).

Therefore, small sample sizes in combination with the lack of representativeness of the target population with which they were designed for, limits both the scientific integrity and generalizability of both the NOMAS and PIBBS to the neonatal population.

Procedures

Whether the examiners rating both the NOMAS and PIBBS were blinded to diagnoses of the patients they were rating was also a significant methodological concern. In three out of the four reviewed investigations (Palmer et al., 1993; Nyqvist et al., 1996; Nyqvist et al., 1999), there was no mention as to whether or not the observers possessed any information about the participants, which could have systematically influenced the way in which they administered, scored, or interpreted the results (Dollaghan, 2007). In addition, it should also be noted for the three aforementioned studies, the creators of the experimental tool also served as observers, which further increases the possibility of experimenter bias.

The time that was given to each rater in both NOMAS studies to assess the recorded material

was not specified and controlled for (Palmer et al., 1993; da Costa & van der Schans, 2008). Although the NOMAS investigations were carried out in such a way as to simulate a real-life clinical experience, the fact that external raters instead viewed recorded sessions at a later date with no mentioned protocol with regards to time or viewing limitations may have affected examiner evaluation. For example, repeated viewings of the recording may have resulted in a feeding skills due to the material.

No standard procedure for time and amount of observations by both examiners and mothers was present in both PIBBS studies (Nyqvist et al., 1996; Nyqvist et al., 1999); assessments took place at any

convenience (Nyqvist et al., 1996), or as often as mothers could (Nyqvist et al., 1999).

Whereas independent raters were noted in both NOMAS studies (Palmer et al., 1993; da Costa & van der Schans, 2008), the PIBBS investigations (Nyqvist et al., 1996; Nyqvist et al., 1999) utilized joint observation sessions between examiners and mothers on several occasions, which may have affected the level of

Official Journal of the California Perinatal Association, 13(1), 28-35.

Rogers, B., & Arvedson, J. C. Assessment of infant oral sensorimotor and swallowing function (2005). *Mental Retardation and Developmental Disabilities Research Reviews, 11(1), 74-82.*

Subramaniam, K. (2001). *A qualitative inquiry into the development and use of knowledge in paediatric occupational therapy.* M.Sc. dissertation, University of Toronto, Canada. Retrieved November 20, 2008, from Dissertations & Theses: Full Text database.

Tooth, L. R., & Ottenbacher, K. J. (2004). The kappa statistic in rehabilitation research: An examination. *Archives of Physical Medicine and Rehabilitation, 85, 1371-1376.*