

# Evaluation of Interrater Reliability for Coding of Types of Gazes in Nurse- Patient Dyads



Alexandria Martz, Student Nurse; Dr. Celia Wills, PhD, RN; Kim Frier, MS, RN, FNP; Dr. Mary Beth Happ, PhD, RN

## Introduction/Background

- Interrater reliability is the amount of agreement between different raters found when assessing the same objects in the same data using the same scale, classification, instrument, or procedure<sup>1</sup>
- Established interrater reliability assessment steps<sup>2</sup> are:
  1. Select one or more appropriate indices
  2. Obtain the necessary tools to calculate the index or indices selected
  3. Select an appropriate minimum acceptable level of reliability for index/indices to be used
  4. Assess reliability informally during coder training (competency)
  5. Assess reliability formally in pilot test
  6. Assess reliability formally during coding of the full sample
  7. Select and follow an appropriate procedure for incorporating the coding of the reliability sample into the full sample coding
  8. Report intercoder reliability in a careful, clear, and detailed manner in all research reports

## Purpose

- Describe and evaluate processes of interrater reliability assessment for coding video recordings through a secondary analysis of nurse-patient interactions in ICU according to four categories of visual gaze
- Analyze ways to improve interrater reliability

## Methods

- Theoretical Framework:
  - Conceptual basis for the analysis of videotape communication was the social theoretical framework of symbolic interactionism<sup>3</sup>
  - Interrater reliability assessment based on the steps described by Lombard et al.<sup>2</sup>
  - Goal of 75% agreement
- Method:
  - Detailed coding criteria and definitions were established for four types of visual gaze
  - Coding agreement of the four types of visual gazes was examined within and across videotapes
  - Raw percentage agreement was calculated by dividing the total number of possible items to be agreed on by the number of times the data collectors agreed

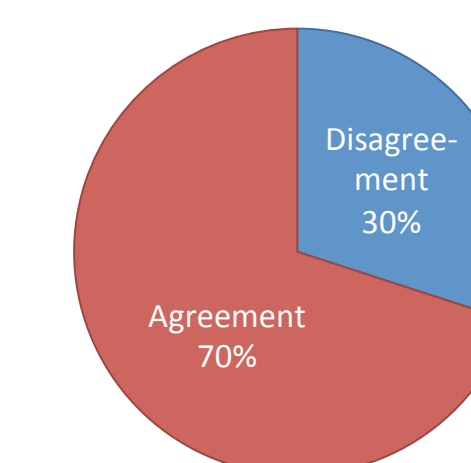
## Sample

- N=13 recordings of ICU nurse-patient dyads collected from a non-intervention sample in a prior study<sup>4</sup>
- Length of each video = 3:15 to 15:27 minutes; mean video length = 6:36 minutes
- Number of gaze occurrences per video = 7 to 29; mean = 18 gaze occurrences

## Results

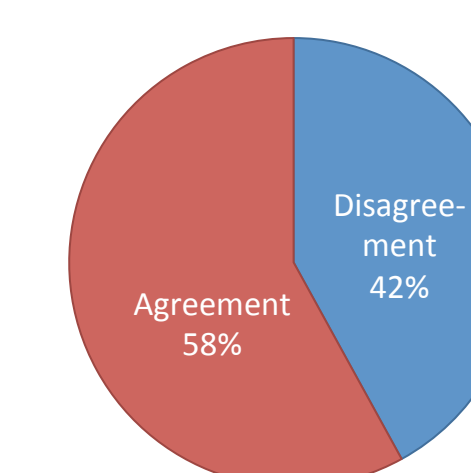
### Interrater Reliability Established for Each Type of Gaze

#### Technical Doing Gaze



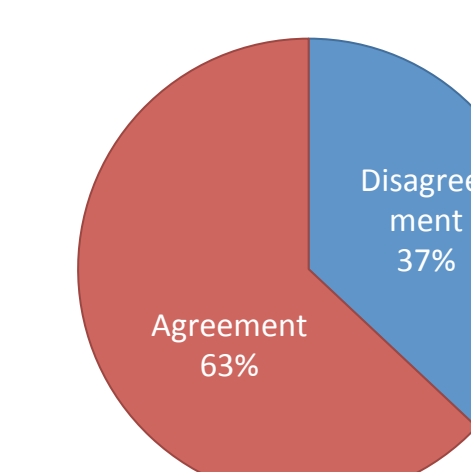
- Disagreement stemmed from confusion between “technical doing” gaze and “assessing” gaze where proxemics needed to be defined

#### Assessing Gaze



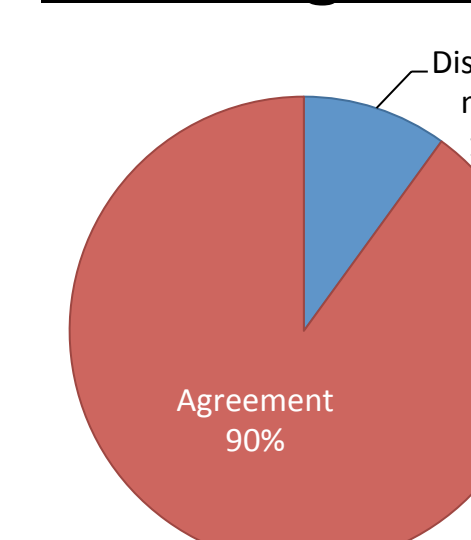
- Coding clarifications were required to distinguish “assessing” gaze from both “technical doing” in its physical distance between the nurse and patient and “listening” in its intent

#### Listening Gaze



- Disagreement for “listening” gaze came from the need to define the reason for the nurse’s gaze in order to distinguish the gaze from “assessing”

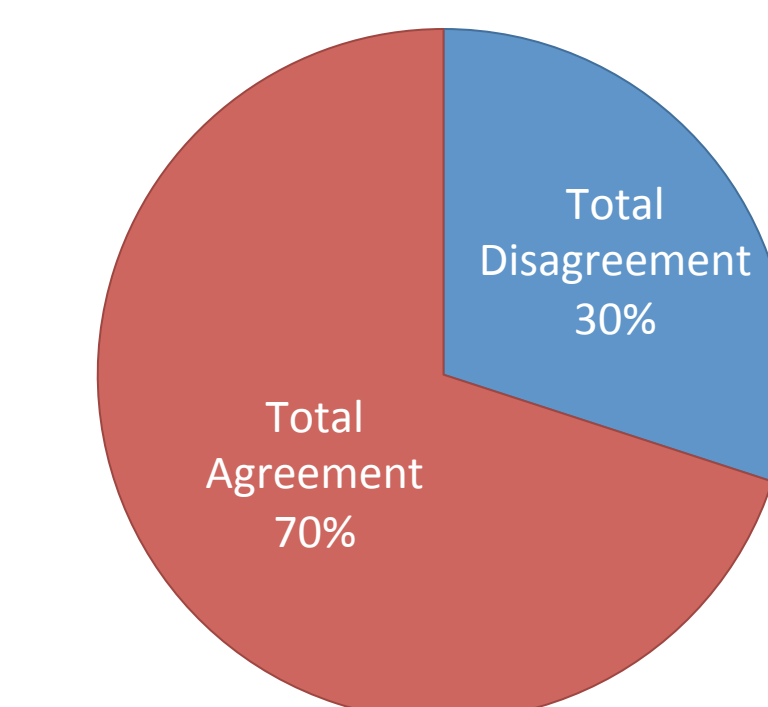
#### Relating Gaze



- The source of disagreement for “relating” gaze was a missed occurrence, rather than a coding disagreement

## Results

### Overall Interrater Reliability



- Overall, coders agreed on 70% of coding indicative of early phase observational scale development
- One gaze, “relating,” achieved the goal of 75% agreement with 90%
- Accuracy of coding increased as codebook definitions were clarified
  - First video: 30% disagreement
  - Final two videos: 22% & 0% disagreement

## Discussion

- To improve raw percentage agreement:
  - Include more practice videos
  - Increase in sample size and length of video
  - Thorough review of codebook definitions prior to coding
- An extension of this project could be to use the Kappa coefficient approach to evaluate IR

## References

- 1 Burns, M. K. (2014). How to establish interrater reliability. *Nursing*, 44(10), 56-58. doi: 10.1097/01.NURSE.0000453705.41413.c6 [doi]
- 2 Lombard, M., Snyder-Duch, J., Bracken, C. C. (2010, June 1). *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects*. Retrieved from <http://matthewlombard.com/reliability/>
- 3 Blumer, Herbert (1969). *Symbolic Interactionism: Perspective and Method*. New Jersey: Prentice-Hall, Inc.
- 4 Happ MB, Garrett KL, Tate JA, DiVirgilio D, Houze MP, Demirci JR, George E, Sereika SM. Effect of a multi-level intervention on nurse-patient communication in the intensive care unit: results of the SPEACS trial. *Heart Lung*. 2014 Mar-Apr;43(2):89-98.

Acknowledgement: Video data from the SPEACS study (National Institute of Child Health & Human Development 5R01 HD043988, M. Happ, PI