Assessing the Non-technical Performance of Teams during Simulated Crises
Psychometric Analysis of Scoring Rubrics

Scott c. Watkins MD, Getulio R de Oliveira Filho PhD, Matthew D. McEvoy MD

INTRODUCTION
- Non-technical skills (teamwork, situational awareness, communication and decision making) are crucial for the management of critical events.
- The purpose of this study was to assess the reliability and scoring variance associated with two tools for measuring the non-technical performance of multidisciplinary perioperative teams.
- Two tools for assessing non-technical skills of teams were identified from the literature, the TEAM tool (Figure 1) and a BARS tool (Figure 2).

METHODS
- We identified two tools for measuring the technical and non-technical performance of teams, trained a group of raters to use the instruments and assessed the reliability of scores from the trained raters.
- Recordings of pediatric perioperative teams managing simulated critical events were reviewed and scored by two group of pediatric anesthesiologist using the TEAM tool and the BARS tool.
- The inter- and intra-rater reliability of scores from trained raters using each instrument was determined.
- G-studies were employed to determine the sources of variance in the scores assigned by the raters using each instrument.
- Decision studies (D Studies) were conducted to estimate the changes in reliability indexes (G-coefficients) secondary to manipulations of the number of different sessions per rater and/or the increase of raters per sessions.

RESULTS
- Within rater reliability (ICC) of the BARS tool ranged from 0.59 to 0.89 and TEAM tool ranged from 0.78 to 0.93 representing good to excellent reliability.
- Between rater reliability (G studies) of the BARS tool ranged from 0.4 to 0.81 and TEAM tool ranged from 0.5 to 0.85 representing acceptable to good reliability.
- Decision studies (D Studies) were conducted to estimate the changes in reliability indexes (G-coefficients) secondary to manipulations of the number of different sessions per rater and/or the increase of raters per sessions.
- The G coefficients of the scores from the TEAM and BARS tools demonstrate excellent reliability and are approaching levels needed for high stakes assessment.
- Variability among raters was responsible for considerable amount of error variance in the scores of non-technical skills instruments, indicating that additional training may be necessary to calibrate raters.
- This study highlights the importance of training professionals in the use of assessment instruments and assessing the reliability and variance of the scores generated by the raters, i.e. assessing the assessor.

TABLES and FIGURES

Figure 1 – BARS Tool

Figure 2 - TEAM Tool

Table 1 - Variance of TEAM and BARS Tools scores

Figure 3. Reliability of BARS measures as a function of the number of raters

Figure 4. Reliability of BARS measures as a function of the number of rating sessions per rater per video

Figure 5. Reliability of TEAM measures as a function of the number of raters

Figure 6. Reliability of TEAM measures as a function of the number of rating sessions

CONCLUSION
- The G coefficients of the scores from the TEAM and BARS tools demonstrate excellent reliability and are approaching levels needed for high stakes assessment.
- Variability among raters was responsible for considerable amount of error variance in the scores of non-technical skills instruments, indicating that additional training may be necessary to calibrate raters.
- This study highlights the importance of training professionals in the use of assessment instruments and assessing the reliability and variance of the scores generated by the raters, i.e. assessing the assessor.