Accuracy Of Assessment Of Team Performance By Team Members In Simulated Intensive Care Crises

Dr Jane Torrie; Dr Associate Professor Jenny Jennifer Weller; Dr Rob Frengley; Dr Boaz Shulruf.
University of Auckland, Auckland, New Zealand
j.torrie@auckland.ac.nz

Aims:
This study compared self-ratings of team performance in simulated intensive care crises, with ratings by expert assessors. Comparison included overall accuracy of self-assessment against external assessors’ ratings, the effect on self-ratings of additional training, and the relationship of overall team performance scores to accuracy of self-rating.

Background:
Effective teamwork is increasingly recognized as important in healthcare outcomes. Accurate self-assessment of teamwork by team members is essential for reflective learning especially in the absence of external skilled feedback. Self-assessment by individuals is known to be poor to moderate but to our knowledge there is very limited information on the assessment of team performance by team individuals within the teammembers.

Methods:
As part of our teamwork research programme into methods of teaching teamwork, 40 intensive care teams, each comprising (1) one doctor and three nurses, managed four different standardised assessment simulated crisis scenarios, two at the beginning and two at the end of a training day (160 scenarios). Scenario order was randomized. Team members rated their team’s technical and behavioural performance on-line immediately after each scenario without discussion. Between the two initial and the two concluding scenarios, teams participated in a training workshop comprising skill-stations, discussions and three training scenarios. No team viewed another team’s scenarios. Facilitated expert debriefing occurred after both training and assessment scenarios. Using the same rating instrument as team members, three expert assessors blinded to order of scenarios rated the assessment four scenarios, generating 160 scenarios with self-ratings by four team members and three expert ratings. The technical rating instrument had 11 items plus a global score (GTS); the behavioural rating instrument had 22 items plus a global score (GBS); a global overall score (GOS) was also required. A seven point anchored Likert scale was used for rating all items.

Correlation coefficients were calculated...

Effect of additional training was assessed by comparing the difference between mean external and self-ratings in the initial scenarios with the difference in the concluding scenarios using...

The relationship of team performance to accuracy of self rating was examined using correlation coefficient... between the mean self-rating and difference between mean self and external ratings...

Results
Considering all 160 scenarios, the mean self-rated GTS, GBS and GOS were moderately correlated with the mean external ratings (correlation coefficients 0.49, 0.57 and 0.49 respectively).

Additional training during the study day did not increase accuracy of global self-ratings...

Comment [Jane1]: Please make this bit correct!
Comment [Jane2]: ditto
Comment [Jane3]: In Jenny’s paper in 2005 Anaesthesia, she used difference vs external rating, rather than difference vs self rating: “Participants whom the judges scored low overrated their performances, while those the examiners scored higher underrated their performances. This relationship is demonstrated in an inverse correlation between the external score and the difference between the self-assessed and external score (rho 0.614; p < 0.0001).
In 18 scenarios, the mean judges’ score was £ 2.5 on the 5-point scale. In 15 (83%) of these low scoring scenarios, the participants awarded themselves a higher score than did the judges. In the 44 higher scoring scenarios, where the judges’ score was > 2.5, only eight (18%) of the participants awarded themselves a higher score than the judges (p < 0.0001).”
Comment [Jane4]: I don’t quite know what to write about this, please help
The correlation between mean self-ratings and the difference between mean self and external ratings was positive and significant for GTS, GBS and GOS (0.29, 0.30, and 0.24 respectively, p<.001). Thus the higher the self-rated score, the larger the difference between external and self-assessment.

Conclusions:

We found that accuracy of self-assessment of team technical and behavioural performance by clinically experienced staff in immersive scenarios was only moderate and did not improve with additional training over a study day. Teams rating themselves highly had the least accuracy. These findings are consistent with the literature on accuracy of individual self-rating and have implications on the educational value of self-reflection by teams without external input.

The effect of profession and duration of clinical experience on self-rating accuracy, including different aspects of teamwork, will be presented at Simtect 2010.

References


