Resuscitation 83 (2012) 1478–1483

Contents lists available at SciVerse ScienceDirect

Resuscitation

journal homepage: www.elsevier.com/locate/resuscitation

Simulation and education

Team performance in resuscitation teams: Comparison and critique of two recently developed scoring tools☆

Anthony McKay[a], Susanna T. Walker[b,*], Stephen J. Brett[c], Charles Vincent[b], Nick Sevdalis[b]

[a] Department of Resuscitation and Outreach, St. Mary's Hospital, Imperial College Healthcare NHS Trust, Praed Street, London W2 1NY, UK
[b] Clinical Safety Research Unit, Department of Surgery and Cancer, Imperial College London, St. Mary's Hospital Campus, 10th Floor QEQM Building, St. Mary's Hospital, Praed Street, London W2 1NY, UK
[c] Center for Perioperative Medicine and Critical Care Research, Department of Anaesthesia and Intensive Care, Hammersmith Hospital, Imperial College Healthcare NHS Trust, Du Cane Road, London W12 0HS, UK

ARTICLE INFO

Article history:
Received 22 February 2012
Received in revised form 6 April 2012
Accepted 22 April 2012

Keywords:
Resuscitation teams
Non-technical skills
Teamworking skills
Adverse events
Patient safety
Assessment tools

ABSTRACT

Background and aim: Following high profile errors resulting in patient harm and attracting negative publicity, the healthcare sector has begun to focus on training non-technical teamworking skills as one way of reducing the rate of adverse events. Within the area of resuscitation, two tools have been developed recently aiming to assess these skills – TEAM and OSCAR. The aims of the study reported here were:

1. To determine the inter-rater reliability of the tools in assessing performance within the context of resuscitation.
2. To correlate scores of the same resuscitation teams episodes using both tools, thereby determining their concurrent validity within the context of resuscitation.
3. To carry out a critique of both tools and establish how best each one may be utilised.

Methods: The study consisted of two phases – reliability assessment; and content comparison, and correlation. Assessments were made by two resuscitation experts, who watched 24 pre-recorded resuscitation simulations, and independently rated team behaviours using both tools. The tools were critically appraised, and correlation between overall score surrogates was assessed.
Results: Both OSCAR and TEAM achieved high levels of inter-rater reliability (in the form of adequate intra-class coefficients) and minor significant differences between Wilcoxon tests. Comparison of the scores from both tools demonstrated a high degree of correlation (and hence concurrent validity). Finally, critique of each tool highlighted differences in length and complexity.
Conclusion: Both OSCAR and TEAM can be used to assess resuscitation teams in a simulated environment, with the tools correlating well with one another. We envisage a role for both tools – with TEAM giving a quick, global assessment of the team, but OSCAR enabling more detailed breakdown of the assessment, facilitating feedback, and identifying areas of weakness for future training.

1. Introduction

In many potentially high-risk industries, like commercial aviation, the nuclear industry, and the oil industry, analyses of human errors have consistently revealed that "human factors", specifically teamworking skills, are often at the heart of errors and failures.[1–3] To reduce human errors and promote safety and high reliability, assessment and training of a range of operators' "non-technical" teamworking skills has been introduced in these industries (often termed "crew resource management" (CRM) training).[4,5] Non-technical skills, including monitoring/situational awareness, decision-making, leadership, and communication skills,[6,7] reflect how operators behave and think during routine activity, but also when crises occur and need to be safely managed.[8]

Following high profile errors resulting in patient harm and attracting negative publicity,[9,10] the healthcare sector as a whole has also turned its attention to non-technical skills – with the specialties of anaesthesia and surgery paving the way. Within these specialties, CRM-styled training has been developed,[11,12] and a range of tools that capture non-technical skills and assess team performance, typically via observation, have been developed and validated for use in real clinical settings as well as in simulation-based training environments.[13–16]

☆ A Spanish translated version of the abstract of this article appears as Appendix in the final online version at http://dx.doi.org/10.1016/j.resuscitation.2012.04.015.

* Corresponding author.
E-mail address: susanna.walker@imperial.ac.uk (S.T. Walker).

0300-9572
http://dx.doi.org/10.1016/j.resuscitation.2012.04.015

Non-technical skills are particularly relevant to resuscitation settings and acutely ill patients.[7,17] When compared with the general hospital population, emergency patient care is especially susceptible to adverse events,[18,19] and Ornato et al[20] have demonstrated that these are associated with decreased survival of adults with in-hospital cardiac arrest. A variety of factors are thought to contribute to this, including time-pressured decision-making, an unstable patient population, an increased number of invasive procedures, and rapid assembly of ad hoc teams. This supports the need for non-technical skills awareness, and training for staff caring for these patients.[21,22] Studies have also shown that effective teamwork may counteract problems with staffing and management, which in itself may reduce the incidence of adverse events.[23,24]

Specifically within the area of resuscitation, two different tools have been developed recently aiming to capture team performance and skills. The first one to be published in the literature was the Team Emergency Assessment Measure (TEAM) (Supplementary Online Appendix A), developed by an Australian research group.[25] TEAM rates 11 behavioural aspects of the whole team on a Likert scale of 0–4, with an additional overall team score rated from 1 to 10. The behaviours that are measured are broken down into Leadership, Teamwork (including communication, co-operation and monitoring/situational awareness), and Task Management.

The second tool is the Observational Skill-based Clinical Assessment tool for Resuscitation (OSCAR) (Supplementary Online Appendix B), developed by our own research group.[26] OSCAR was based on a rating tool previously developed and extensively validated for use in operating theatre settings, called the Observational Teamwork Assessment for Surgery (OTAS).[14,15] OSCAR rates the performance of individual sub-teams within a standard resuscitation team (anaesthetists, physicians and nurses) across six teamwork-related behaviours (communication, co-operation, co-ordination, monitoring/situational awareness, leadership, and decision-making). Examples of "ideal" behaviours are given for each team member in each behaviour mode category to assist the assessor in determining ability. Each behaviour is rated on a 0–6 Likert scale with an additional "overall" score given for each section.

Another tool was developed almost concurrently by Andersen et al. in Denmark.[27] However, this rates entire team behaviours on a dichotomous ("yes" and "no") scale in a checklist format. The distinction between technical and non-technical performance within it is not as clear as it is within either OSCAR or TEAM. Due to these differences in skill content and coverage, we chose not to include this third tool in the direct comparison.

TEAM and OSCAR have been developed independently but are similar in their aim to capture team processes and performance. The aim of this study was to compare psychometrically the two tools, and determine the overall validity of the skills-assessment that is quantified within the context of resuscitation. Psychometric comparison in tools that involve observational assessment should include statistical evaluation of inter-rater reliability – in other words, the level of agreement between assessors using the tools. High reliability indicates that a tool produces consistent results across different assessors.[28] Therefore the first two research questions that we addressed were:

What is the inter-rater reliability of OSCAR?

What is the inter-rater reliability of TEAM?

Moreover, given that OSCAR and TEAM aims to capture very similar skill sets, albeit in subtly different ways, we also directly compared assessments of the same resuscitation teams carried out using each one of the two tools. This is a question of concurrent validity, which addresses whether two instruments designed to assess similar skills and behaviours actually produce comparable assessments when used concurrently.[28] Our final research question, therefore, was:

To what extent do OSCAR and TEAM scores correlate (i.e., statistically measure similar team characteristics)?

2. Methods

2.1. Procedure

2.1.1. Phase 1 – reliability assessment

This phase aimed to assess the inter-rater reliability of both tools to ensure that they can be used reliably in assessing team skills in resuscitation contexts. Reliability assessment was performed by watching 24 pre-recorded resuscitation simulations (Supplementary Online Appendix C). The simulations had all been performed by cardiac arrest teams from our hospital (teaching hospital, London, UK). Twenty took place within the hospital's simulation centre, with small resuscitation teams consisting of a physician, an anaesthetist, and two nurses. These lasted an average of 5.5 min each. Four additional simulations were carried out "in situ" in clinical areas of the hospital, performed by the real on-duty resuscitation team for the day. These were inevitably longer simulations, and lasted an average of 13.5 min each.

The simulation recordings were watched by two resuscitation experts; one resuscitation officer (AMcK), and one anaesthetist (SW). Assessors were kept blinded to each other's ratings throughout this phase, and were trained in their observations prior to the beginning of the study. Each assessor watched each video once and applied both tools (i.e., OSCAR and TEAM).

2.1.2. Phase 2 – content comparison and correlation of scorings

In this phase, the structure and use of the two tools were critically compared, and then the team ratings they generated were statistically correlated and plotted. Strong positive correlations between the two tools would provide evidence that they are broadly quantifying the same skill-sets (i.e., evidence for concurrent validity).

2.2. Statistical analyses

Data analyses were carried out using SPSS v.18.0 (SPSS Inc., Chicago, IL, USA). Inter-rater reliability refers to the level of agreement between two (or more) assessors using an assessment instrument. Intraclass correlation coefficients (ICC) were used to assess this in both TEAM and OSCAR, as recommended in the literature – with ICC values of 0.70 or higher indicating adequate agreement in scoring.[22] Moreover, we also carried out non-parametric Wilcoxon tests to test whether the average scores allocated by each assessor were significantly different (non-significant results would indicate the desirable consistency in the scoring between the two assessors).[14] Concurrent validity was assessed using non-parametric Spearman's rho correlation coefficients between OSCAR and TEAM scores. Scatterplots of these correlations, as well as Bland–Altman plots were produced. Bland–Altman plots are typically used to assess the level of agreement between two different measurement tools.[29]

Given the differences in the structure of the two tools, some algebraic manipulation was necessary for the correlational analyses to be possible. We computed an average score on each tool and expressed it as a percentage score (%). For the TEAM tool, we based the analyses on the first 11 questions, which are all scored on 0–4 point scales. The final question that assesses global performance on a 10-point scale was not included in this analysis, as it is scored on a different scale, it does not assess an individual skill or behaviour, and there is no OSCAR equivalent for comparison. TEAM scores, potentially ranging between 0 and 44, were then expressed as a percentage. For the OSCAR tool, there are six behaviours scored separately for three subgroups (anaesthetists, physicians, and