ORIGINAL RESEARCH

Assessing human factors during simulation: The development and preliminary validation of the rescue assessment tool

John Unsworth, Andrew Melling, Jaden Allan, Guy Tucker, Michael Kelleher

Department of Healthcare, Faculty of Health & Life Sciences, Northumbria University, UK.

Correspondence: John Unsworth. Address: Northumbria University, Coach Lane, Newcastle-upon-Tyne, England, UK. Email: john.unsworth@northumbria.ac.uk

Received: December 28, 2013 **Accepted:** February 8, 2014 **Online Published:** March 11, 2014
DOI: 10.5430/jnep.v4n5p52 **URL:** http://dx.doi.org/10.5430/jnep.v4n5p52

Abstract

Background: Failure to rescue the deteriorating patient is a concern for all healthcare providers. In response to this problem providers have introduced a range of interventions to promote timely rescue. Human factors and non-technical skills play a part in both the recognition of ill patients and in the delivery of interventions associated with their successful rescue. Given the risks to patient safety which failure to rescue raises, simulation provides a vehicle for staff training and development in terms of both technical and non-technical skills. This paper describes the development and preliminary validation of a human factors rating tool specifically designed to assess the non-technical skills associated with the recognition and rescue of the deteriorating patient.

Methods: Using high fidelity simulation scenarios related to patient deterioration Faculty independently rated student performance. Scoring took place using video footage of the students' performance. Data were analyzed to establish the validity of the tool, internal consistency between categories and elements and inter-rater reliability.

Results: Content validity was established through a process of review and by checking for duplicate or redundant items. The internal consistency of the tool was acceptable with a Cronbach's alpha of 0.84. Factor analysis suggested that the tool assessed only two components rather than the three hypothesized during tool development. The components were labelled as recognizing and responding and leading and reassuring. Inter-rater reliability was initially poor at 0.21 but following training of raters this rose to above 0.8 for two videos related to the same scenario one which had been used during training. However, when the scenario changed the reliability dropped to 0.5.

Conclusions: Rescue appears to be a well-structured tool with good levels of inter-rater reliability following intensive training related to the specific scenario being scored. Further work is required to establish all aspects of construct validity and to ensure test-retest reliability.

Key words

Deterioration, Rescue, Non-technical skills, Assessment

1 Introduction

Failure to rescue deteriorating patients is a global concern [1-3] which relates to the failure to prevent patient deterioration arising from a complication of an underlying illness [4]. Failure to rescue has become a significant issue because of the

increased acuity of patients on general wards [5]. Nurses are often well placed to detect the early signs and symptoms of possible complications and increased vigilance makes timely rescue more likely [1]. However, the recognition and rescue of a deteriorating patient is a complex process and there are numerous points for potential failure including not taking or recording observations, a failure to appreciate the significance of the results of observations or a failure to respond appropriately and communicate concern to other practitioners [6]. This suggests that the appropriate recognition and rescue of the deteriorating patient relies upon a mixture of technical and non-technical skills. A number of authors have identified how nursing staffing levels and more recently the nurses' level of skills and knowledge can have an impact on the recognition and rescue of deteriorating patients [6-8].

In response to the problem of a failing to rescue in the UK the National Institute for Clinical Excellence (NICE) issued guidance around the recognition of acute illness in hospital patients [9]. Across many western health systems the utilization of physiological track and trigger systems [10] and the establishment of critical care outreach teams [11] have been implemented in order to reduce the likelihood of failure to rescue. While the use of track and trigger systems may improve recognition of deterioration there is still the possibility that human factors could play a part in communicating concern to other practitioner and ensuring an appropriate response to the patient's deterioration. The term human factors, has been defined as 'the scientific discipline concerned with the understanding of interactions amongst humans and other elements of the system' [12]. In essence human factors theory operates on two levels to influence the system or the individual. At a systems level, human factors theory seeks to influence the design of processes, tasks, equipment and the workplace to make allowances for human capabilities when working in the environment. As such human factors at a systems level seek to design out error or institute processes which are designed to compensate for the existence of human imperfection [13]. At an individual level human factors theory seeks to describe and enhance the non-technical skills which compliment an individual's technical skills to promote safe and effective task performance [14]. Human factors have been implicated in around 80% of all patient safety incidents [15] and given the complex processes involved in the recognition and rescue of the deteriorating patient it is likely that several non-technical skills are involved including decision making, team working, leadership and situational awareness [16].

Simulation techniques have been widely utilized to educate nurses and other health professionals about the recognition of the deteriorating patient [17]. Most studies have utilized high fidelity simulation with reflective review and de-brief. While such approaches report high participant satisfaction and improvements in self-reported knowledge and confidence they also highlight how in some cases participants made errors with missed 'cues' and incomplete assessment [18]. Many simulation programs [19, 20] based around the deteriorating patient incorporate feedback to participants about non-technical skills the measurement of performance around such skills is less common. One such study [21] examined situational awareness amongst final year nursing students during high fidelity simulation scenarios. While students were able to identify physiological indicators of deterioration during the scenarios their ability to integrate this with other knowledge about possible reasons for the changes and to develop comprehension of the situation was lower in both the hypovolemic shock and sepsis scenarios.

Flin and colleagues [14] outline how the assessment of non-technical skills may be carried out to:

- Provide feedback to trainees on skill development
- Test competence
- Ascertain whether a non-technical skills training program has been effective
- Audit the level of skill demonstrated by a work team

While behavioral rating scales may be useful in assessing non-technical skills they are not without their limitations [14]. These limitations include the fact that they cannot capture all aspects of performance as they are limited to assessing what can be observed. Therefore, important non-technical skills such as decision making, coping with fatigue and managing

stress are less amenable to observation than other skills. The scales are unlikely to be able to assess infrequent events unless these can be recreated using simulation.

A number of non-technical rating scales have been developed for use within healthcare. These scales are used to provide feedback to practitioners on their non-technical skills either through simulated practice or following the observation of actual practice. NOTSS (Non-Technical Skills for Surgeons) [22] is a behavioral rating scale which is used to assess surgeons' non-technical skills in the areas of decision making, communication and teamwork, task management, situational awareness and leadership. The tool was initially developed and subject to an evaluation of its validity and reliability by asking surgeons to rate each, others, performance during six simulated operating room scenarios. The initial evaluation indicated good internal consistency between the constructs being assessed and satisfactory levels of agreement in terms of average measures. A subsequent evaluation [23] highlighted the need for training for raters as this revealed how novice raters were likely to rate the performance lower than those individuals who were regarded as experts. Acceptable levels of inter-rater reliability were found for two categories on the NOTSS behavioral rating scale and high levels of consistency using average measures. The authors go on to acknowledge that absolute agreement is the most appropriate model. Despite this the absolute agreement was only at an acceptable level for 2 out of the 20 categories in the scale.

A behavioral marker system which was designed to assess non-technical skills amongst anesthetists (ANTS – Anaesthetists Non-Technical Skills) was developed in 2003 [24]. The tools sought to assess similar non-technical skills to NOTSS and the initial development and psychometric testing of the tool involved anesthetists rating each, others, performance during anesthetic simulation scenarios. The authors indicate that ANTS has internal consistency between the constructs being assessed and, that satisfactory levels of agreement and accuracy were obtained between raters. The authors go on to outline how training of raters is important so that the tool can be rolled out into anesthetic training programs. Reasonable levels of inter-rater reliability were achieved despite the large sample size and the rater's lack of familiarity with the ANTS scale. However, the level of agreement was around 0.5. The authors acknowledge that with intensive training levels of 0.7 and above may be achievable in raters with an awareness of human factors.

The CATS (Communication and Team Skills) assessment tool [25] is an instrument for measuring team performance around non-technical skills in both a practice setting and during simulation. The instrument rates four domains using a three point scale ranging from observed and good, variation in quality and expected but not observed. The instrument was trialed but was not subjected to comprehensive evaluation to assess its psychometric properties.

Given the central role which nurses' play in the recognition and rescue of the deteriorating patient there is a need to ensure that registered nurses have both the technical and non-technical skills to ensure patient safety from the point of registration. This paper describes the development of *Rescue* a human factors rating scale which was designed to assess the non-technical skills of under-graduate nurses during simulation scenarios. Within our under-graduate programs all nursing students undertake a range of simulation sessions all of which address technical and non-technical skills required to recognize and rescue the deteriorating patient. During the course students' progress through a range of simulation sessions which are designed to complement the taught components of the course. The taught components include teaching the students about medical and surgical complications, the management of specific conditions such as sepsis and teaching providing the theory behind human factors and non-technical skills. All of the simulation sessions include a preparation session designed to prepare the students for the scenario and to outline the aims and objectives of the session and a structured debrief immediately following the scenario. Feedback is provided on both technical performance and on non-technical skills.

This paper specifically examines the development of the *Rescue* rating scale and the preliminary validation of the tool and the assessment of inter-rater reliability when the tool is used to rate student performance.

2 Subjects and methods

The methods used in the preliminary evaluation of the tool aimed to commence the process of validation and to identify whether the tool produces results which are reliable across a range of raters. Validation of the tool will provide an indication of the extent to which the assessment tool measures a particular construct in a particular context [26]. There is no single measure or metric for measuring whether a tool is valid and while validity is described as encompassing content, criterion and construct validity it is now widely acknowledged that construct validity acts as single unitary concept [27]. A series of pre-specified hypotheses were developed prior to data collection [28]. Table 1 details the evaluation criteria, hypotheses, data sources, analysis methods and a summary of the results for the preliminary validation and testing of the *Rescue* tool.

Table 1. Evaluation criteria

Evaluation criteria	Hypotheses	Data source	Analysis method	Summary of results
Validity	RESCUE provides a comprehensive set of categories and elements to describe the non-technical skills associated with the recognition and rescue of the deteriorating patient	Review by experts Identification of missing, duplicated or redundant items	Review by authors and critical review by other experts	
Reliability	Users of RESCUE will achieve inter-rater agreement and consistency at acceptable levels	Ratings data related to consistency and absolute agreement	Inter-Class Correlation	**ICC-2 Consistency** ***Video 1*** – 0.84 (95% confidence interval) 0.60 0.95 ***Video 2*** – 0.89 (95% confidence interval) 0.74 0.97 ***Video 3*** – 0.58 (95% confidence interval) -0.05 0.89 **ICC-2 Absolute agreement** ***Video 1*** – 0.84 (95% confidence interval) 0.60 0.95 ***Video 2*** – 0.86 (95% confidence interval) 0.65 0.96 ***Video 3*** – 0.52 (95% confidence interval) -0.34 0.86
Internal consistency	RESCUE will have an acceptable level of internal consistency between categories and levels	Ratings data	Cronbach's alpha	0.84
Tool structure	RESCUE has a structure based around three components – communication; leadership and situational awareness; assessment and problem solving	Ratings data	Factor analysis	Kaiser-Meyer-Olkin co-efficient 0.86 Bartlet test of sphericity $x^2 = 368.70, df = 36, p \leq .00$ Component 1 – recognizing and responding Component 2 – leading and reassuring

2.1 Developing the tool and rating scale

To identify which constructs should be included in any rating scale the development team identified that any successful rescue involves at least two phases; surveillance and the timely identification of problems and the subsequent response. In any deteriorating patient scenario this would require the practitioner to elicit information through communication and

physical assessment, identify a problem, develop and appraise options, prioritize and delegate as appropriate and then communicate with colleagues in order to ensure that the patient's condition is responded to. As highlighted earlier not all non-technical skills are amenable to observation of behaviors and performance. Having identified a number of the elements involved in the recognition and rescue of the deteriorating patient the team were able to identify the broad categories which the tool should include. These were communication, leadership, situational awareness, problem solving and decision making and workload management. An initial version of the *Rescue* tool was developed using these broad categories and the elements shown in Table 2. The development team identified that each of the elements included in the tool were addressed during the simulation sessions and at various points as part of lectures and seminars within the program.

Table 2. Initial RESCUE tool categories and elements

Category	Elements
Communication	Communicates with the patient to keep them informed of what is happening and to offer reassurance Nurse to Nurse / Doctor communication – gives clear concise information and instructions Handover includes: S – situation B – background A – assessment R – recommendation(s)
Leadership	Identifies leader who assumes and continues in role throughout Co-ordinates team using authority / assertiveness as necessary
Problem solving	Problem identification Data gathering / Analysis Identification of risks Development and appraisal of options
Situational Awareness	Immediate awareness of situation and / or patient's condition Anticipates possible sequale
Workload Management	Prioritization – appropriate identification of immediate and short term priorities Team or Leader assigns roles clearly Leader appropriately delegates in terms of risk, experience and the task

A four point rating scale was selected for the *Rescue* tool. The ratings were poor, marginal, good and outstanding. Each point on the scale would have a behavioral marker to indicate expected performance at each level for the element being assessed. An example of the behavioral anchors for one of the element of the *Rescue* tool can be found at Figure 1 Assessment scales with behavioral anchors are commonly used when assessing non-technical skills [23, 24]. However, there is little evidence that behaviorally anchored scales are more superior to other forms of rating [29].

A four point scale was selected because it was perceived as easy to use and because similar four point scales had been used to develop rating scales in aviation [30] and for the non-technical rating scales developed to date for use with doctors [23, 24].

The development team debated whether the *Rescue* tool should be used as a criterion or norm referenced assessment. As a criterion referenced assessment students would be assessed against the standard of a registered nurse irrespective of the stage they were at in their undergraduate program. Most assessments in UK Higher Education are norm referenced [31] where students are judged against the standard for a student at their stage of education. However, there is increasing recognition that nursing and other practice based professions need to move towards criterion referenced assessed to ensure that practitioners are fit for practice at the point of qualification [32]. Eventually, the development team agreed that *Rescue* should be used as a criterion referenced assessment tool.

Figure 1. The *Rescue* Assessment tool

CATEGORY	ELEMENTS	GOOD	ACCEPTABLE	MARGINAL	POOR
Communication	Communicates with the patient to keep them informed of what is happening and to offer reassurance	Communicates with the patient throughout to explain what is going on and to reassure ◯	Provides some explanation to the patient in order to gain co-operation and provide reassurance ◯	Minimal communication with the patient. Verbal communication restricted to gathering data / information ◯	Little or no effective communication with the patient with no essential information being gathered ◯
	Professional to professional communication – gives clear concise information and instructions using SBAR as appropriate for handovers	Communicates with colleagues in a structured, clear and concise way to provide information / instructions ◯	Communicates with colleagues to provide instructions ◯	Communicates but instructions / messages are not concise but do not affect the outcome ◯	Fails to adequately communicate with colleagues which leads to errors in patient management ◯
Leadership/ Workload Management	Identifies leader who assumes and continues in role throughout	Clear identification of a leader who leads and co-ordinates the team throughout ◯	Clear identification of a leader who leads and co-ordinates the team for the majority of the time ◯	No identification of a leader but a team member assumes role with some attempts at team co-ordination ◯	Poor team co-ordination with no leader identified or assumed ◯
	Prioritization – appropriate identification of immediate and short term priorities	Elicits information and identifies immediate and subsequent short term priorities ◯	Elicits information and identifies immediate priorities but no planning of other short term priorities ◯	Elicits information and identifies some immediate priorities although actions are slightly out of sequence ◯	Elicits information but fails to prioritize actions resulting in time delays ◯
	Leader appropriately delegates in terms of risk, experience and the task	Clear delegation between team members with tasks appropriate to risk, knowledge and experience ◯	Delegation between team members of tasks with some attempts to ensure task is appropriate ◯	Delegation between team members with little reference to the complexity of the task being delegated ◯	Inappropriate or no delegation to team members of tasks ◯
Problem solving /Situational Awareness	Problem identification	Early identification of cause for concern using EWS or vital obs / patient presentation ◯	Timely identification of cause for concern before further patient deterioration ◯	Identification of cause for concern after further patient deterioration ◯	Failure to identify cause for concern ◯
	Immediate awareness of situation and / or patient's condition	Clear assessment of patient using all relevant items of A B C D E approach. Assessment of situation from handover ◯	Assessment using some of the relevant elements of the A B C D E approach to identify cause for concern ◯	Recognition of cause for concern using handover or initial presentation. Assessment lacks systematic focus ◯	Failure to recognize cause for concern and risks from situation ◯
	Development and appraisal of options	Identifies an appropriate range of options based on patient presentation. Takes action in relation to one of these options ◯	Identifies a single option based on patient presentation and takes action ◯	Identifies an option but no evidence of a rationale for this judgment ◯	Fails to identify an option or take action without a clear rationale for this decision ◯
	Anticipates possible problems which may follow initial presentation (sequale)	Aware of possible sequale and observes for these to ensure early intervention to prevent further deterioration ◯	Some awareness of possible sequale and observation for these ◯	Ensures observation of the patient's condition and takes appropriate action ◯	Fails to recognize possible sequale and therefore defaults to routine observation frequencies ◯

Once the initial version of the tool had been developed it was subject to review by the development team to identify duplicate and potentially redundant items. In addition, reviewers were asked to ensure that there was sufficient differentiation between each of the behavioral anchors and that the performance cited in each of the four points of the scale would equate to poor, marginal, good or outstanding performance from a registered nurse. Following the review a number of redundant and duplicate items were removed and several categories were amalgamated to avoid repetition. The final structure of *Rescue* was hypothesized to be based around three categories; communication leadership and workload management and problem solving and situational awareness. Within these three categories there were nine elements which would be rated. These were:

- Communication with the patient
- Professional to professional communication
- Identification of a leader
- Identification of immediate and short term priorities
- Delegation from the leader to others
- Problem identification
- Immediate awareness of the patient's condition
- Development and appraisal of options
- Anticipation of possible problems

2.2 Testing the tool

Two simulation scenarios were selected for the initial evaluation of the *Rescue* tool. The simulation scenarios are delivered in the second year of a three year undergraduate nursing degree program. Both scenarios relate to the recognition and rescue of a deteriorating patient. One scenario is based around an elderly female patient admitted to hospital with a urinary tract infection. The patient goes on to develop a bacteremia and sepsis. The second scenario involves a male patient with known ischemic heart disease admitted with chest pain. The pain worsens despite the administration of a nitrate. The simulation scenarios are of high fidelity and Laedral SimMan™ or Sim essential™ mannequins are used during scenarios.

2.3 Pilot of scoring

A number of authors [23, 24] have cited how the training of raters is paramount if the scores obtained from a measure are to be regarded as reliable. It is not clear how much training is required especially when the ratings are being carried out by experienced Faculty members. In order to ascertain the training requirements an initial pilot scoring exercise was undertaken. An initial team of 12 raters were identified from Faculty staff all of who were involved in facilitating simulation with undergraduate nurses. Raters were given a short 45 minute session to introduce them to the tool and to discuss ratings. During the tool familiarization session participants were shown a video example of good performance and of poor performance although they were not asked to rate these videos. Following this the raters independently scored the performance of four groups of students undertaking either a sepsis or chest pain scenario by observing video recordings of the sessions. Following the scoring pilot the data were analyzed using the Statistical Package for the Social Sciences (SPSS). This work was undertaken to assess the inter-rater reliability of the scoring. Inter-rater reliability refers to the degree to which different raters give consistent estimates to the same phenomenon. To assess inter-rater reliability an Intra-class Correlation (ICC -2) was performed to compare the scores between raters. ICC assesses the reliability between raters by comparing the variability of different ratings of the same subject to the variation across all ratings and all subjects. An ICC is reported as a score ranging from 0 (no agreement) to 1 (absolute agreement).

 ISSN 1925-4040 E-ISSN 1925-4059

2.4 Pilot results

The ICC for the four pilot scoring exercises was 0.21 in terms of absolute agreement and 0.87 for consistency. While the score for consistency is at an acceptable level the score for absolute agreement was poor. As the *Rescue* tool is a criterion based assessment, absolute agreement is important as this relates to whether individual raters score the same phenomenon in the same way rather than whether they produce scores which enable ranking of participants. The pilot scores for reliability suggested that more intensive and comprehensive training was needed to improve inter-rater reliability.

2.5 Training and scoring

Following the pilot the development team worked on designing a program of more specific rater training. It was decided that all raters should be trained at the same time so that the results obtained during training could be discussed and debated. Six raters were given three hours of intensive training during which they independently rated and then discussed the scores given to five videos related to the identification and management of sepsis and chest pain. Elements of each *Rescue* score were discussed in detail until convergence was reached in subsequent ratings. Immediately following the training raters rated three videos of student performances (two related to a sepsis scenario and one to chest pain). Ratings were completed in real time and there was no discussion. Raters were given five minutes to complete their scoring between each video. Each rater recorded a unique numerical identifier to ensure that their ratings were anonymized.

2.6 Statistical analysis

Within the pilot and subsequent scoring sessions each element within the *Rescue* tool was assigned a numerical score to each behavioral marker. Therefore, not observed was scored as a 0, Poor as a 1, Marginal as a 2, Good as a 3 and Outstanding as a 4. Data from each rater were entered into the Statistical Package for the Social Sciences (SPSS). An Intra Class Correlation (ICC-2) was then calculated for overall consistency and to assess absolute agreement. Within ICC values of 0.81 to 1.0 denote almost perfect agreement, 0.61 go 0.80 denote substantial agreement and 0.10 – 0.6 denote slight to moderate agreement [33].

The internal consistency of the tool was assessed using Cronbach's alpha [34]. Cronbach's alpha is expressed as a value between 0 and 1 and values of 0.8 and above are regarded as having good internal consistency. At the same time where a test or tool seeks to measure more than one construct e.g. is not uni-dimensional, as is the case with *Rescue*, then Factor Analysis should be considered [35].

Exploratory factor analysis was conducted to test the hypothesized structure of the tool. Principal component analysis as used as the method of factor extraction. Factor analysis is a method of data reduction which is used to determine the number of factors a scale measures based on a particular number of inter-related quantitative variables. Factor analysis can be used in scale development to identify redundant and duplicate items and to examine the dimensions of unobservable entities [36].

3 Results

Internal consistency of the tool was assessed using Cronbach's alpha and a result of 0.84 was obtained suggesting good internal consistency between the items on the test.

The reliability of the tool was assessed by a series of intra-class correlations, examining both consistency and absolute agreement. For consistency the results show 0.84 (95% confidence interval 0.60 0.95) for video one, 0.89 (95% confidence interval 0.74 0.97) for video two and 0.58 (-0.05 .89) for video three. Absolute agreement results ranged from 0.84 (95% confidence interval 0.60 0.95) for video one, 0.86 (95% confidence interval 0.65 0.96) for video two to 0.52 (95% confidence interval -0.34 0.86) for video three.

The factor analysis revealed that the Kaiser-Meyer-Olkin coefficient was 0.86 indicating good sampling adequacy and the Bartlet test of sphericity was statistically significant ($x^2 = 368.70$, df = 36, $p \leq .00$) indicating that the correlation matrix justifying the carrying out of the factor analysis. Varimax rotation revealed two components (see Table 3) which accounted for 66.31% of the variance. While the hypothesized structure of the tool was not confirmed by the factor analysis all of the variables were loaded onto the two components identified. The variables associated with communication with patients and with professionals each loaded on to separate components. The factor analysis revealed no redundant or duplicate items and all of the variables were accounted for in the two components identified through principal component analysis.

Table 3. Factor Analysis – Components Identified via Principal Component Analysis

Component One	Component Two
Label assigned: recognizing and responding	**Label assigned:** leading and reassuring
Variables:	**Variables:**
Communication with professionals	Communication with the patient
Problem identification	Leader identified
Immediate awareness of condition	Prioritization
Option appraisal	Delegation
Anticipate problems	

Table 3 details the two components and the variables loaded onto each. Component One was identified as containing the variables most closely associated with recognizing and responding with immediate awareness of condition, option appraisal, anticipation and problem solving sitting alongside communication with other professionals.

Component Two was identified as containing the variables associated with leading and reassuring and this included leadership of the team, prioritization and delegation alongside communication with the patient.

4 Discussion

The results suggest that *Rescue* has a consistent internal structure measuring non-technical skills broadly grouped into categories relating to leading and communicating and recognizing and responding. There is an appropriate level of internal consistency between items on the tool. Following comprehensive training of raters acceptable scores for absolute agreement between rater were achieved in two intra-class correlations and the rating for one student group fell below acceptable levels at 0.15. The disagreement between raters appeared to be associated with a move from norm referenced assessment e.g. judging students against the expectations for their stage within a nurse education program toward criterion referenced assessment. The scores in a number of areas were higher amongst some raters who also provided feedback about whether the expectations of the students were appropriate for those who were in the second year of a three year under-graduate program. The transition between norm referenced and criterion referenced assessment has been acknowledged as problematic for academics that are used to the well-established norm referenced methods of assessment used within UK Higher Education [11]. Part of the problem is the link between criterion referenced assessment and the assessment of competence [37] and the fact that many academics consider that students should not be regarded as competent until they are near the end of their program of study. There are also concerns that allowing students to mirror the performance of a registered nurse could result in the student believing that they are now allowed to practice the new skill without supervision.

In addition, the lower score for video three may be associated with training as moderate to good levels of agreement had been achieved when rating performance in the scenario around sepsis which had also been used as the scenario for training purposes. Video three related to chest pain suggesting the training specific to the scenario being rated may be appropriate rather than simply training raters to use the tool.

The approach adopted in the evaluation of the *Rescue* tool was to evaluate non-technical skills alongside the performance of technical skills. This means that in any de-brief students would receive feedback about their technical performance as well as about non-technical human factors. This is the approach which was adopted by the development team in the delivery of simulation session prior to the development of *Rescue*. Providing feedback about non-technical skills alongside technical performance has a number of advantages including the fact that the non-technical skills are contextualized as part of the overall approach adopted by students in recognizing and responding to patient deterioration. The team plan to continue this approach in the further development of the tool. Although it may be possible to conduct simulation sessions which purely relate to non-technical skills it would appear to be inappropriate to separate out specific aspects of student performance given the complex interplay between technical and non-technical skills.

Strengths and limitations

One of the strengths of the *Rescue* tool is that it is fairly easy to complete and the results are then immediately available. This is particular important given that the tool has been designed to use during simulation sessions. The availability of staff resource solely for scoring may be limited as simulation is a resource intensive activity. This means that the either the Faculty facilitating the session or the member of staff playing the patient could also be asked to score the student's performance using the *Rescue* tool. Such timely feedback can be valuable to the student as it allows them to contextualize the scores and immediately work to improve performance in subsequent simulation sessions.

One of the limitations of the current study is that it only includes a preliminary analysis of the assessment tools validity and reliability. Further work in required to test convergent validity and such work can be achieved using similar scoring tools developed for the assessment of non-technical skills amongst other professional groups. In addition, the current study only explored a small number of student groups, all of who were in the second year of their under-graduate nursing program, and this prevented any assessment of whether the tool has predictive validity. Further work is also needed to establish whether the tool has intra-observer and test retest reliability that is, do observers score students the same when the tool is used at different times.

5 Conclusion

Non-technical skills, alongside technical skills play an important role in the recognition and rescue of the deteriorating patient. The development of non-technical skills amongst under-graduate nurses is vital if they are to be deemed fit for practice at the point of qualification. In order to promote non-technical skill development formative assessment of such skills during simulation should be commonplace. The *Rescue* assessment tool has been specifically developed to enable the rapid assessment of non-technical skills during simulation scenarios. The results of the assessment can be then be immediately fed back to students to enable them to identify areas of good practice as well as those areas which require improvement.

In order to promote reliable assessment assessors need training in the use of the Rescue tool. Such training should be scenario specific and should be based on the use of criterion referenced assessment.

Further work to evaluate Rescue is required to establish test retest reliability and to identify if the tool has predictive validity. Despite the need for further work Rescue appears to allow for rapid assessment of, and feedback on, student performance during simulation sessions.

References

[1] Clarke S P, Aiken, L H: Failure to rescue. American Journal of Nursing 2003; 103: 42-47. http://dx.doi.org/10.1097/00000446-200301000-00020

[2] Jones S, Bottle A and Griffiths P: An assessment of "failure to rescue" derived from routine NHS data as a nursing sensitive patient safety indicator for surgical inpatient care, 2011; National Nursing Research Unit, Kings College: London

[3] Buykx P, Kinsman, L, Cooper S, McConnell-Henry T, Cant R, Endacott R et al: FIRST2ACT: Educating nurses to identify patient deterioration – a theory based model for best practice simulation education. Nurse Education Today. 2011; 31: 687-693. http://dx.doi.org/10.1016/j.nedt.2011.03.006

[4] Silber JH, Williams SV, Krakauer H, and Schwartz JS: Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Medical Care. 1992; 30: 615-629. http://dx.doi.org/10.1097/00005650-199207000-00004

[5] Haines S and Coad S. Supporting ward staff in acute areas: expanding the service. Intensive and Critical Care Nursing. 2001; 17: 105-109. http://dx.doi.org/10.1054/iccn.2001.1563

[6] Luettel D, Beaumont K, Healey F: Recognising and responding appropriately to the early signs of deterioration in hospital patients. 2007; National Patient Safety Agency: London

[7] Clarke SP: Balancing staffing and safety. Nursing Management. 2003; 34: 44-48. http://dx.doi.org/10.1097/00006247-200306000-00018

[8] Aiken LH, Clarke SP, Cheung RB, Sloane DM and Silber JH: Education levels of hospital nurses and surgical patient mortality. Journal of the American Medical Association: 2003; 290: 1617-1623. http://dx.doi.org/10.1001/jama.290.12.1617

[9] National Institute for Clinical Excellence: Acutely ill patients in hospital: recognition of and response to acute illness in adult in hospital. 2007; NICE: London

[10] Jansen JO, Cuthbertson BH: Detecting critical illness outside of the ICU: the role of track and trigger systems. Current Opinion in Critical Care 16. 2010; 184-190. http://dx.doi.org/10.1097/MCC.0b013e328338844e

[11] McArthur-Rouse F: Critical care outreach services and early warning scoring systems: a review of the literature. Journal of Advanced Nursing. 2001; 36: 696-704. http://dx.doi.org/10.1046/j.1365-2648.2001.02020.x

[12] Carayon P (ed) Handbook of human factors and ergonomics in healthcare and patient safety. 2007; CRC Press Danvers MA.

[13] Youngberg BJ: Patient safety handbook (2nd ed) Jones and Bartlett Publishers, Burlington MA (pp157-168)

[14] Flin R, O'Connor P and Crichton M: Safety at the Sharp End: a guide to non-technical skills. Ashgate: Surrey (pp1-16)

[15] NHS Education for Scotland: Patient Safety and Clinical Skills: Human Factors. Available from: http://www.nes.scot.nhs.uk/education-and-training/by-theme-initiative/patient-safety-and-clinical-skills/introduction-to-patient-safety/human-factors.aspx - [accessed 18 November 2012]

[16] National Outreach Forum: Critical Care Outreach 2003: progress in developing services. Department of Health NHS Modernisation Agency: London

[17] LeBlanc VR, Manser T, Weinger MB, Musson D, Kutzin J and Howard SK: The study of factors affecting human and system performance in healthcare using simulation. Simulation in Healthcare. 2011; 6: S24-S29. http://dx.doi.org/10.1097/SIH.0b013e318229f5c8

[18] Endacott R, Scholes J, Cooper S, McConnell-Henry T, Porter J, Missen K, et al: Identifying patient deterioration: using simulation and reflective interviewing to examine decision making in a rural hospital. International Journal of Nursing Studies. 2012; 49: 710-717. http://dx.doi.org/10.1016/j.ijnurstu.2011.11.018

[19] Kim J, Neilpovitz D, Cardinal P, Chiu M and Chinch J: A pilot study using high fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: the University of Ottawa critical care medicine high fidelity simulation and crisis resource management 1 study. Critical Care Medicine. 2006; 34: 2167-2174. http://dx.doi.org/10.1097/01.CCM.0000229877.45125.CC

[20] Yee B, Naik VN, Joo HS, Savoldelli GL, Chung DY, Houston PL, et al: Non-technical skills in anesthesia crisis management with repeated exposure to simulation based education. Anesthesiology. 2005; 103: 241-248. http://dx.doi.org/10.1097/00000542-200508000-00006

[21] Cooper S, Kinsman L, Buykx P, McConnell-Henry T, Endacott R and Scholes J; Managing the deteriorating patient in a simulation environment: nursing students' knowledge skill and situational awareness. Journal of Clinical Nursing. 2010; 19: 2309-2318. http://dx.doi.org/10.1111/j.1365-2702.2009.03164.x

[22] Yule S, Flin R, Paterson-Brown S, Maran N, and Rowley D. Development of a rating system for surgeons' non-technical skills. Medical Education. 2006; 40: 1098-1104. http://dx.doi.org/10.1111/j.1365-2929.2006.02610.x

[23] Yule S, Flin R, Maran N, Rowley D, Youngerson G and Patterson-Brown S: Surgeons' Non-technical Skills in the Operating Room: reliability testing of the NOTSS behavior rating system. World Journal of Surgery. 2008; 32: 548-556. http://dx.doi.org/10.1007/s00268-007-9320-z

[24] Fletcher G, McGeorge P, Flin R, Galvin R and Maran N: Anaesthetists' non-technical skills (ANTS): evaluation of a behavioral marker system. British Journal of Anaesthesia. 2003; 90: 580-588. http://dx.doi.org/10.1093/bja/aeg112

[25] Frankel A, Gardner R, Maynard L and Kelly A: Using the Communication and Teamwork Skills (CATS) assessment to measure healthcare team performance. Joint Commission Journal on Quality and Patient Safety. 2007; 33: 549-558.
[26] Sechrest L: Validity of measures is no simple matter. Health Services Research. 2005; 40: 1584-1609. http://dx.doi.org/10.1111/j.1475-6773.2005.00443.x
[27] Moss PA: Reconstructing validity. Educational Researcher. 2007; 36: 470-476. http://dx.doi.org/10.3102/0013189X07311608
[28] Guyatt GH, Feeny DH, Patrick DL: Measuring health related quality of life. Annuals of Internal Medicine. 1993; 118: 622-629. http://dx.doi.org/10.7326/0003-4819-118-8-199304150-00009
[29] Schwab DP, Heneman HG, DeCotiis TA: Behaviourally anchored rating scales: a review of the literature. Personal Psychology. 1975; 28: 549-562. http://dx.doi.org/10.1111/j.1744-6570.1975.tb01392.x
[30] Helmieich RL and Merriot AC: Culture at work in aviation and in medicine. Aldershott: Ashgate
[31] Dunn L, Parry S and Morgan C: Seeking quality in criterion referenced assessment. Education-Line. Available from: http://www.leeds.ac.uk/educol/documents/00002257.htm [accessed 08/10/2012]
[32] Redfern S, Norman S, Calman L, Watson R and Murrells T: Assessing competence to practice in nursing: a review of the literature. Research Papers in Nursing Education. 2002; 17: 51-77. http://dx.doi.org/10.1080/02671520110058714
[33] Kramer MS, Feinstein AR: Clinical biostatistics. LIV. The biostatistics of concordance. Clinical Pharmacological Therapy. 1981; 29: 111-123. http://dx.doi.org/10.1038/clpt.1981.18
[34] Cronbach L: Coefficient alpha and the internal structure of tests. Psychomerkia. 191; 16: 297-334.
[35] Tavakol M and Dennick R: Making sense of Cronbach's alpha. International Journal of Medical Education. 2011; 2: 53-55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd
[36] Streiner DL and Norman GR: Health measurement scales: a practical guide to their development and use. 2008; (4th edition); Oxford: Oxford University Press (pp 96) http://dx.doi.org/10.1093/acprof:oso/9780199231881.001.0001
[37] Morgan C, O'Reilly M: Assessing open and distance learners. 1999; London: Kogan Page.