Rating Anaesthetists’ Non-Technical Skills – The ANTS system

R. Flin\textsuperscript{1}, G. Fletcher\textsuperscript{1}, P. McGeorge\textsuperscript{1}, R. Glavin\textsuperscript{2}, N. Maran\textsuperscript{2} & R. Patey\textsuperscript{3}

Paper presented at 47\textsuperscript{th} Human Factors and Ergonomics Society Conference,
Denver 2003

\textsuperscript{1}\text{Department of Psychology, University of Aberdeen, King’s College, Aberdeen, AB24 2UB, SCOTLAND}
\textsuperscript{2}\text{Scottish Clinical Simulation Centre, Stirling Royal Infirmary, Livilands Gate, Stirling, FK8 2AU, SCOTLAND}
\textsuperscript{3}\text{Department of Anaesthesia, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, AB25 2ZN, SCOTLAND}

Corresponding author: Rhona Flin, Department of Psychology, University of Aberdeen, King’s College, Old Aberdeen AB242UB, SCOTLAND. e-mail: r.flin@abdn.ac.uk
Abstract

Non-technical skills are the cognitive and social skills required in any operational task involving decision making and team work, such as aviation or surgical medicine. This study identified the key non-technical skills required in anaesthesia, and developed a behavioural marker system (Anaesthetists’ Non-Technical Skills (ANTS)) for their assessment in theatre, in the simulator or from video recordings. It comprises four skill categories (task management, team working, situation awareness and decision making) divided into fifteen elements, each with example behaviours. A first stage evaluation was conducted by asking 50 consultant anaesthetists to rate the behaviour of a target anaesthetist using the prototype ANTS system in eight videos of simulated anaesthetic scenarios. The system appeared to be complete, and the skills were observable, and could be rated with acceptable levels of agreement and accuracy. The second stage evaluation is now testing the usability of the behavioural rating system in the operating theatre environment.
Summary

Background.
Non-technical skills are the cognitive and social skills required in any operational task involving decision making and teamwork, such as aviation or surgical medicine. These are often referred to as Crew Resource Management (CRM) skills. In European aviation, taxonomies of non-technical skills with associated behavioural marker sets and rating forms have been developed for the training and assessment of pilots’ non-technical skills (Flin & Martin, 2001; Flin et al, under review, O’Connor et al 2002).

Similar systems are now being developed for training, assessment and research within the acute medical specialities (Byrne & Greaves, 2001; Gaba et al, 1994, 1998; Greaves & Grant, 2000; Helmreich et al, 1995). Non-technical skills provide a critical underpinning for good practice but are not always addressed explicitly in standard medical training (Fletcher et al 2001). Realisation of the need to train and assess these skills is growing but these activities must be based on properly developed skill frameworks, validated measurement tools and adequate rater training (Baker et al 2001; Holt et al, 2001; Klampfer, Flin, Helmreich et al 2001). The aim of this study was to identify the key non-technical skills required by anaesthetists (anesthesiologists) and to develop and evaluate a behavioural marker system for their assessment in theatre, in the simulator or from video recordings.

Method. Based on previous methods used in aviation (Flin et al in press; O’Connor et al 2002) a prototype system was devised. This occurred in four stages: (i) literature review (see Fletcher et al 2001); (ii) examination of existing marker systems (see above); (iii) interviews using cognitive task analysis techniques (Flanagan, 1954; Klein et al 1989; Seamster et al, 1997, 2001; Strauss & Corbin, 1990) (iv) workshops and an iterative development process. The resulting Anaesthetists’ Non-Technical Skills (ANTS) System comprises four skill categories (task management, team working, situation awareness and decision making – see Table 1) divided into fifteen elements, each with example behaviours.
A formal evaluation was then conducted by asking 50 consultant anaesthetists to rate the behaviour of a target anaesthetist using the prototype ANTS system in eight video taped scenarios. These were filmed in the Scottish Clinical Simulation Centre and depicted a range of anaesthetic scenarios, such as interactions with other team members e.g. surgeons, critical situations, routine problem solving, task handover. A group of Subject Matter Experts (consultant anaesthetists) previously rated the behaviours displayed in each scenario to provide a set of benchmark or reference ratings.

**Results.** Evaluation results from the videotape study (see Table 2, 3, 4) indicate that the system appears to be complete, the skills are observable, and can be rated with acceptable levels of agreement and accuracy (Fletcher et al in press). The system has good internal consistency, and responses regarding usability were very positive.

The current phase of research involves testing the usability of the system in the operating theatre environment. In this evaluation, consultant anaesthetists are trained to use the ANTS rating system and then they collect ratings of trainee anaesthetists when they are working in the operating theatre. This is normal training practice for postgraduate specialists in medicine, they work under the supervision of consultants who provide ongoing training and regular feedback on their performance. Multi-centre ethical approval has been granted to permit the rating system to be used by anaesthetists during normal operating theatre lists in five hospitals. Details of the ANTS system will be presented with results of the first videotape evaluation system and the second phase of usability trials. The suitability of this skills taxonomy and behavioural marker rating system for training (see Maran et al 2002; Pizzi et al 2001), assessment and research purposes will be discussed.

**Acknowledgements**
The ANTS System was developed under funding from the Scottish Council for Postgraduate Medical and Dental Education, now part of NHS Education for Scotland, from September 1999 – August 2003. The views presented in this paper are those of the authors and should not be taken to represent the position or policy of the funding body. We would like to extend our grateful thanks to all the anaesthetists who have taken part in the research and the organisations that have supported this.
References


Greaves JD, Grant J. Watching anaesthetists’ work: using the professional judgement of consultants to assess the developing clinical competence of trainees. Br J Anaesth, 2000: 84; 525-533.


Williams DM, Holt RW, Boehm-Davis DA. Training for inter-rater reliability: baselines and benchmarks. *Proceedings of the 9th International Symposium on Aviation Psychology; 1997 Apr 27-May 1; Columbus, Ohio.* *Columbus: Ohio State University*
Table 1. ANTS System – Category level Definitions

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Management:</td>
<td>Managing resources and organising tasks to achieve goals, be they individual case plans or longer term scheduling issues.</td>
</tr>
<tr>
<td>Team Working:</td>
<td>Working with others in a team context, in any role, to ensure effective joint task completion and team satisfaction; its focus is particularly on the team rather than the task.</td>
</tr>
<tr>
<td>Situation Awareness:</td>
<td>Developing and maintaining an overall dynamic awareness of the situation based on perceiving the elements of the theatre environment: patient, team, time, displays, equipment, understanding what they mean and thinking ahead about what could happen in the near future.</td>
</tr>
<tr>
<td>Decision Making:</td>
<td>Making decisions to reach a judgement or diagnosis about a situation, or to select a course of action, based on experience or new information under both normal conditions and in time-pressured crisis situations.</td>
</tr>
</tbody>
</table>
**Table 2. Summary of results for validity**

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Results</th>
</tr>
</thead>
</table>
| Completeness of ANTS System          | 1) Did it address the key NTS behaviours displayed? **Yes = 100%;** *(n=50).*  
2) Do you think and elements or categories are missing? **No = 84%;**  
Yes = 8%; Comment only = 8%; *(n=50).*  
3) Do you think any elements or categories are superfluous? **No = 81% ;** Yes = 17%; comment only = 2%; *(n=46)* |
| Observability of NTS                 | 1) Averaged across all scenarios NTS observability was good, ranging from 100% (gathering info, recognising & understanding) to 66% (assessing capabilities). Overall, 13 elements observable >80%, and all categories were observable >95%.  
Across all scenarios, only 5% of ratings (8 out of 152) showed no difference between use of ‘observed’ vs ‘not observed’, Chi-squared not significant.  
2a) How easy was it to relate behaviours to elements? **Average or Easy = 78%;** Difficult or Very difficult = 22%; *(n=50)*  
2b) How easy was it to relate behaviours to categories? **Very easy to Average = 82%;** Difficult or Very difficult = 18%; *(n=50).*  
Categories were significantly easier to relate to behaviours than elements *(t=-3.06, p<0.05)* |

**Table 3. Summary of results for reliability**

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-rater Agreement</td>
<td>Levels of <strong>$r_{wg} = 0.55-0.67$</strong> at element level, with ‘recognising &amp; understanding’ being lowest and ‘identifying 7 utilising resources’ being highest; and at category level <strong>$r_{wg} = 0.56-0.65$</strong>, with Situation Awareness being lowest and Task Management and Team Working being joint highest. Overall Situation Awareness and its individual elements were associated with the lowest levels of agreement. Values of $r_{wg}$ varied across scenarios.</td>
</tr>
<tr>
<td>Rater against Reference ratings</td>
<td>‘Accuracy’ as measured by rater-reference rater difference: &gt; <strong>88% accuracy to one scale point.</strong> Mean absolute deviation (MAD) from the reference i.e. an error score = <strong>0.49-0.84</strong>, which while showing significant difference between elements again suggests only minor differences arising across boundaries.</td>
</tr>
<tr>
<td>Internal Consistency</td>
<td>Correlations were strongest indicating the ‘best-fit’ for existing category-element grouping for 13 of the 15 elements; for remaining two mapping was not the highest on 2-3 scenarios. Consistency between elements in each category, Cronbach alpha ranged from <strong>0.79-0.86</strong>, suggesting commonality but not duplication.</td>
</tr>
</tbody>
</table>
Table 4. Summary of results for usability

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Results</th>
</tr>
</thead>
</table>
| **Acceptability**   | 1) Was the ANTS System useful for structuring observations? **Yes = 100%**  
2) Would the ANTS System be helpful for consultants giving training to junior anaesthetists? **Yes = 94%** *(describes skills/gives framework, useful for problem trainees)*; comment only=6% *(more evidence & familiarity needed)*  
3) Would the ANTS System be helpful for consultants assessing junior anaesthetists? **Yes = 78%** *(gives it structure/puts it into words; some with caveats e.g. training, validation)*; No=8% *(informal feedback/self-assessment; not good enough at assessment)*; comment only=7% *(validation, more familiarity practice)*  
4) Do you think the ANTS System could be used to support in-theatre teaching? **Yes = 94%** *(highlights important skills, help with observation/giving feedback)*; No=4%; comment only=2% |
| **Design of ANTS System** | 1) Was the wording used for category and element labels meaningful? **Yes = 98%**; No = 2%; *(n=48)*  
2) Were the descriptions clear? **Yes = 96%**; No = 4% ; *(n=48)*  
3) Were the ‘good’ behavioural markers helpful? **Yes = 96%**; comment only = 4%; *(n=50)*  
4) Were the ‘poor’ behavioural markers helpful? **Yes = 92%**; comment only = 8%; *(n=48)*  
5) Do you think the rating scale gave you enough flexibility to rate the performance levels see? **Yes = 94%**; No = 6%; *(n=50)*  
*Comments varied across both groups with some preferring a longer scale, some a shorter, some wanting a mid-point and others not.* |