Original Article

Cognitive Aids in Medicine Assessment Tool (CMAT): preliminary validation of a novel tool for the assessment of emergency cognitive aids*

D. Evans,[1] R. McCahon,[2,3] M. Barley,[3] A. Norris,[2,3] A. Khajuria[4] and I. Moppett[5]

1 Airway Fellow, 2 Honorary Lecturer, 4 Medical Student, 5 Associate Professor, Anaesthesia and Critical Care Research Group, University of Nottingham, Nottingham, UK
3 Consultant, Department of Anaesthesia, Queen's Medical Centre, Nottingham, UK

Summary

Applying human factors principles to the design of clinical emergency guidelines is important. The UK Civil Aviation Authority uses a Checklist Assessment Tool for evaluating the content and usability of emergency drills before introduction into service on aircraft. We hypothesised that this model could be used to develop a generic medical tool. A three-stage modified Delphi process was used to adapt the above tool for use in designing medical emergency guidelines. The resulting Cognitive aids in Medicine Assessment Tool was then used to score and rank seven published difficult airway guidelines; the scores were used to assess its validity and reliability. Pearson's rank coefficient between these scores and scores from independent assessors was 0.89 ($p = 0.007$). Internal consistency, as assessed by Cronbach's alpha, was 0.74, 0.96 and 0.72 for the tool's three constituent domains of physical characteristics, content and layout/format, respectively. Inter-rater reliability, as assessed by Cohen's kappa, ranged from 0.33 to 0.72. The adoption of our tool has the potential to improve the usability of medical emergency guidelines.

Correspondence to: D. Evans
Email: daveevans@doctors.net.uk
Accepted: 27 December 2014
**Presented in part at the Difficult Airway Society Annual Scientific Meeting, Ascot, UK; November 2013.*

Introduction

Checklists, drills and guidelines are increasingly used in medicine to address the well-recognised problems of poor performance and flaws in individuals' recall. The logic and content of many cognitive aids reflect expert opinion, but their physical characteristics, layout and format appear to receive less attention and may make them poorly suited for use in an emergency.

The increasing availability of new methods of diagnosis and treatment has helped to create a challenge to modern medicine: namely, how to deliver known best practice reliably and with minimal harm. Clinical emergencies can involve multiple factors that must be taken into account in a short period, creating problems with cognitive overload. Checklists and guidelines designed for these scenarios can be grouped together under the term 'cognitive aids'. Such aids appear to be beneficial; there is evidence demonstrating reduced mortality with checklists for central venous access on intensive care units [1] and surgical safety checklists in

operating theatres [2–4]. The probability of demonstrating that cognitive aids reduce morbidity or mortality in emergency care is limited by the rarity of these events; however, in the field of anaesthesia and surgery, checklists have been shown to improve the management of simulated intra-operative crises [5, 6]. Much of the design process for pre-existing cognitive aids has focused on content; a recent review of 22 cognitive aids used in anaesthetic practice found only one that employed any human factors design methodology [7].

The use of cognitive aids is a long-standing feature of aviation and other safety-critical industries. In the late 1980s and early 1990s, several major airline accidents were attributed to the misuse of checklists; as a result, guidelines were proposed to promote the use of ergonomic principles in the design of checklists [8]. Design factors associated with aviation checklist failure and non-compliance include: typography; format; layout; organisation; wording; level of detail; and logical coherence [9]. A widely used tool for aviation checklist assessment is the UK Civil Aviation Authority's Checklist Assessment Tool (CHAT) [10]. It was developed in partnership with airlines and consultants from the human factors industry and has been in use, with two revisions, since 1997. This Checklist Assessment Tool has been praised by researchers from the US National Aeronautics and Space Administration as 'the most complete document(s) pertaining to checklists that are currently available' [11].

We hypothesised that this aviation tool could be adapted and validated to assess medical emergency cognitive aids. Our underlying intention was to develop a tool that could be used during the design of new cognitive aids. We used the field of difficult airway management, as there are a number of published aids within this area of practice.

Methods

The study comprised three stages. Stage 1 developed a new tool for assessing medical cognitive aids, which we named the Cognitive aids in Medicine Assessment Tool (CMAT). Stage 2 assessed the validity and reliability of this new tool by using it to assess cognitive aids that have been published for use in difficult airway scenarios. Stage 3 assessed the utility of the new tool outside its original testing domain.

Stage 1: Development of the prototype CMAT

The original aviation tool comprises 72 desirable attributes and associated descriptors in three domains of physical characteristics, content and layout/format. We employed a modified Delphi approach [12], drawing on our own expertise as anaesthetists with an academic interest in human factors, to develop the new medical tool.

In the first Delphi round, each of the five authors independently placed each of the 72 attributes and descriptors found in the aviation Checklist Assessment Tool into one of the following categories:

A Attribute should be retained. Initial inclusion was based on achieving agreement from at least four of the five authors.

B Attribute should be retained, but requires alteration to conform to a medical model.

C Attribute should be rejected, as this attribute has no relevance to a medical model. Initial rejection was based on achieving agreement from at least four of the five authors.

Attributes with agreement from fewer than four of the five authors were allocated to category B for further discussion. In the second Delphi round, category B attributes and descriptors were revised by one of the authors (DE) and circulated to the other authors for independent comment. The category B attributes were then revised, and combined with those from category A, to form a first draft. In order to retain the validity of the well-established aviation tool, the format was kept as close as possible to the original. A meeting of all five authors was then held to discuss and, where necessary, revise agreed attributes and descriptors, to produce a second draft. We devised the following scoring system for our tool. If an item completely adhered to the attribute described, it was scored as 2; if it partially adhered to the attribute it was scored as 1; and if it did not adhere to the attribute it was scored as zero. It was also possible to judge an attribute as 'not applicable'. After this, the authors met for a final time to discuss the results and agree a final draft.

Stage 2: Assessment of the CMAT using cognitive aids for difficult airway management

The intended use for our new tool is to aid the design of new cognitive aids. In order to assess its reliability and validity as an assessment, the five authors who developed the tool in Stage 1 used CMAT to evaluate a sample of difficult airway cognitive aids. These were found by forward and backward searching through Google and Google Scholar, as well as the author's knowledge, and revealed over 30 documents. From this list, seven published documents were chosen that had either been produced or endorsed by a national body or society (category 1), or that showed evidence of a validation process or the use of a research methodology (category 2) (Table 1). Each document was allocated at random to two of the five authors, so that it was scored independently according to the rating scale outlined above. Paper documents were printed from online content in A4 size and in colour, unless otherwise stated by the original authors, before being allocated. These original scores were used for the inter-rater agreement analysis described below. Each investigator was blinded to the other's scores; the two scores were summed to give a single score for each of the tool's 43 attributes. If an attribute was scored as 'not applicable' by one author, but the second gave it a score, then these authors were asked to revise their individual scores. If this failed to resolve the disagreement then a final decision on 'not applicable' scores was made by DE. The maximum overall score a cognitive aid could achieve was 172 (43 attributes with a maximum score of two from both investigators). As different cognitive aids have varying proportions of attributes that were not applicable, the final score is displayed as fraction of the maximum achievable score to avoid bias against shorter documents with less content.

We assessed the reliability of this process in two ways. First, to assess the internal consistency we calculated Cronbach's alpha coefficient [20], from the scores generated by the authors in Stage 2, for each of the three substantive domains within the CMAT construct: physical characteristics; content; and layout/format. Cronbach's alpha is not robust against missing data, so attributes with more than two out of seven 'not applicable' scores were excluded from analysis. We calculated the standard error of measurement for the scores generated for each of the three domains described above [21]. To gauge inter-rater agreement between the two authors' scores for each aid, we calculated the chance-adjusted inter-rater agreement (with 95% CI), between pairs of investigators, using Cohen's kappa coefficient [22].

The validity of our tool was assessed by how well it performed in scoring the seven cognitive aids selected in Table 1. There was no existing standard for assessing the scores generated, so we decided to compare CMAT's rank scores with a ranked global impression of usability and quality. To prevent bias, we asked a group of four different anaesthetists specialising in difficult airway management, employed in our hospital but blinded to the purpose of our work, to provide the global impression ranking. Correlation between these two groups was assessed by the two-tailed Pearson's rank coefficient. Statistical analysis was performed throughout with SPSS (IBM SPSS Statistics for Windows, Version 21.0, Released 2012; IBM Corp., Armonk, NY, USA).

Table 1 Difficult airway cognitive aids assessed by the Cognitive Aids in Medicine Assessment Tool.

Cognitive aid	Category	Description
AIM Stanford [13]	2	Developed by Anesthesia Infomatics and Media Lab, Stanford University School of Medicine
Amathieu et al. [14]	2	Published
The Anaesthetic Crisis Manual [15]	1	Endorsed by the European Society of Anaesthesiology
Apfelbaum et al. [16]	1	Published by American Society of Anesthesiologists
Henderson et al. [17]	1	Published by the UK Difficult Airway Society
projectcheck.org [18]	2	Developed by Ariadne Labs, a joint centre for health systems innovation at Brigham and Women's Hospital and Harvard School of Public Health
Stephens et al. [19]	2	Published

Stage 3: Assessment of non-anaesthetic cognitive aids

In order to test its applicability outside the domain of difficult airway management and anaesthetic practice, the five authors who developed the tool in Stage 1 used the new tool to assess three further cognitive aids used in emergency care outside the field of anaesthesia (Table 2). These aids represent a range of clinical urgency, from paediatric resuscitation (where the user would be expecting to read commands aloud to other members of a team) to the slightly less urgent scenario of life threatening asthma (where users might be expected to read the document themselves).

Results

Stage 1: Development of the prototype CMAT

From the 72 attributes described in the original aviation tool, two were unanimously rejected in the first Delphi round, and 34 were initially retained; these were universally context-neutral and related to physical format, layout and text differentiation. The remaining 36 attributes requiring revision were predominantly more specific to an aviation context, though medical parallels were clear; these were moved into the second Delphi round. Further revision and rejection of attributes took place in the second and third Delphi rounds, and resulted in a final draft, containing the 43 attributes (21 attributes identical with the original tool and 22 modified attributes) shown in the Appendix. It contains the same three domains: physical characteristics; content; and layout and format, but with a reduced number of attributes in each section.

Table 2 Non-anaesthetic cognitive aids assessed by the Cognitive Aids in Medicine Assessment Tool.

Cognitive aid	Description
Paediatric advanced life support [23]	Published by Resuscitation Council (UK) in 2010
Shoulder dystocia [24]	Published by the Royal College of Obstetricians and Gynaecologists (UK) in 2012
Severe acute asthma [25]	Published by British Thoracic Society and Scottish Intercollegiate Guidelines Network (revised) 2012

Stage 2: Assessment of the CMAT using cognitive aids for difficult airway management

Overall scores using our new tool to assess seven difficult airway cognitive aids are shown ranked in terms of percentages of their maximum score in Table 3. The only attributes that had universal scores of 'not applicable' from all five authors were those describing 'memory items'. Memory items are commonly used by the aviation industry and are defined by the Civil Aviation Authority as 'actions normally resulting from an emergency situation which must be performed immediately by the crew without reference to any checklist' [10]. The ranking of the cognitive aids from our independent airway experts is also shown. Pearson's coefficient of the independent global rank with the rank score from CMAT was 0.89 ($p = 0.007$). Internal consistency, as measured by Cronbach's alpha, for each of the three domains was 0.74 (standard error of measurement 5.3%) for physical characteristics, 0.96 (standard error of measurement 2.1%) for content and 0.72 (standard error of measurement 2.7%) for layout and format. Inter-rater agreement, as measured by Cohen's kappa, between two authors scoring the same guideline, is shown in Table 4.

Stage 3: Assessment of non-anaesthetic cognitive aids

The results from the assessment of three non-anaesthetic cognitive aids are displayed in Table 5. The clinical specialities of the authors meant they were unable to assess attribute 3.6, 'Does the order of the action items ensure return to a safe state at the earliest opportunity?' and so this was removed from analysis in Stage 3. Inter-rater agreement and overall scores were similar to that achieved with the seven difficult airway cognitive aids.

Discussion

We found that the original assessment tool developed by the Civil Aviation Authority could be adapted successfully to a medical model. The resultant CMAT provided a valid and reliable assessment tool of difficult airway cognitive aids, based on good-to-excellent internal consistency, fair-to-substantial inter-rater reliability and high levels of agreement with independent, expert assessment. In a recent editorial in this journal

Table 3 Comparison of ranked Cognitive aids in Medicine Assessment Tool scores and independent assessor scores for seven difficult airway cognitive aids.

Cognitive aid	Attributes scored (maximum 43)	Maximum score achievable	Actual score	% of maximum achievable score	CMAT ranking	Independent expert ranking
AIM Stanford [13]	36	144	112	77.8	1	3
Amathieu et al. [14]	30	120	31	25.8	7	7
The Anaesthetic Crisis Manual [15]	36	144	84	58.3	4	4
Apfelbaum et al. [16]	30	120	37	30.8	6	6
Henderson et al. [17]	34	136	85	62.5	3	2
projectcheck.org [18]	28	112	83	74.1	2	1
Stephens et al. [19]	29	116	52	44.8	5	5

Table 4 Inter-rater agreement for the seven difficult airway cognitive aids assessed by the Cognitive aids in Medicine Assessment Tool as measured by Cohen's kappa.

Cognitive aid	Cohen's kappa (95% CI)	Inter-rater agreement
AIM Stanford [13]	0.51 (0.72–0.30)	Moderate
Amathieu et al. [14]	0.33 (0.51–0.15)	Fair
The Anaesthetic Crisis Manual [15]	0.51 (0.69–0.32)	Moderate
Apfelbaum et al. [16]	0.52 (0.71–0.33)	Moderate
Henderson et al. [17]	0.47 (0.67–0.26)	Moderate
projectcheck.org [18]	0.72 (0.88–0.55)	Substantial
Shephens et al. [19]	0.71 (0.87–0.55)	Substantial

Table 5 Cohen's kappa and percentage of maximum achievable score for three non-anaesthetic cognitive aids assessed by the Cognitive aids in Medicine Assessment Tool.

Cognitive aid	Cohen's kappa	CMAT score (% of maximum)
Paediatric advanced life support [23]	0.58	62.9
Shoulder dystocia [24]	0.65	51.8
Severe acute asthma [25]	0.49	59.5

[26], Jenkins asserted that design of cognitive aids is important, and that poor design can lead to patient harm; he proposed the development of new aids for the management of anaesthetic room emergencies. We believe that our tool is the only one currently available to guide the design of future cognitive aids.

Validity refers to the 'evidence presented to support or refute the meaning or interpretation assigned to assessment results' [27]. Reliability of assessment refers to the consistency of measurement, and is essential evidence of validity [28]. We assessed the reliability of the CMAT for assessing difficult airway cognitive aids by measuring both the internal consistency and inter-rater agreement. Internal consistency is the extent to which the items in an assessment measure the same construct; a Cronbach's alpha of 0.70–0.95 is generally considered desirable [29]. Overall, the internal consistency was excellent for content, and good for physical characteristics and layout and format. The standard error of measurement for Cronbach's alpha for each of the domains was similar to that previously reported in the medical literature [30].

Inter-rater agreement was found to be substantial for two difficult airway cognitive aids, moderate for four and fair for one [31]. The 95% CI show that the study was underpowered with respect to inter-rater correlation. Only two of the seven cognitive aids assessed had a lower 95% CI above the minimum level of 0.40 suggested by Sim and Wright [32]. We can propose two reasons for the slightly disappointing levels of inter-rater agreement, which reduce our tool's reliability. Firstly, it is likely to reflect the subjective nature of the material assessed; further refinement of attributes would be necessary to ensure consistent interpretation by assessors. We did consider a fourth Delphi round, where attributes would be designed with

clearer interpretation and to make 'not applicable' scores less likely, using a broader range of experts in psychology, engineering and human factors. This was not pursued at this stage, as further revision may have reduced the content of CMAT and decreased the inherent validity of using the original aviation tool as a template. Secondly, it may reflect inadequate rater calibration in the use of the tool. Further work might involve development of assessment item 'anchors' for the description of attributes (e.g. 'this attribute should be scored 1 for the following reasons…') to help calibrate the scoring by different assessors. Again, this would have meant further deviation away from the successful aviation model.

It could be argued that the sample size we used for our modified Delphi technique was small, consisting of the five authors, and from a single site. However, there appears to be no universally accepted minimum panel size for the Delphi method and published work in the field of medicine has used a similarly small number of experts [33, 34]. Furthermore, such a small size may not affect the validity of our work. A study by Akins and colleagues considered the panel size used in the Delphi technique and concluded: 'Experts who have similar training and general understanding in the field of interest allow for effective and reliable utilisation of a small sample from a limited number of experts' [35].

We sought further evidence of validity by comparing the ranking from our new tool with a ranked global score from an independent group of airway experts. As in the modified Delphi technique, the sample size was small and from a single site. Overall correlation between the two was very strongly positive, suggesting that the same underlying attribute was indeed being measured. The UK Difficult Airway Society guidelines [17] were ranked higher by the group of experts and this may represent the familiarity that UK anaesthetists have with these.

A deeper examination of scoring patterns generated by our new tool revealed no explicit examples of memory items in any of the cognitive aids assessed. In most medical emergencies, the initial care provided to the patient will be recalled from memory before the relevant cognitive aid can be found and used. We would suggest that the relevance of memory items to emergency airway management (and emergency medical care in general) is clear. Future cognitive aids, and medical training in general, should consider the applicability of memory items to their areas of practice. Difficult airway aids appear to score universally well on attributes concerning size of document, font type, minimising steps required to complete and order of actions. Two of these attributes relate to logic and this may reflect the consensus of expert opinion that formed them. Attributes that scored poorly on the CMAT were checklist objective, explicit statement of responsibility for actions, deferred items, end-of-drill indication and inclusion of a statement of currency. Designers may wish to consider these deficits in the design process of future cognitive aids.

This work suggests that the new tool can be used to assess cognitive aids outside its initial area of validation. Agreement between the investigators was similar for cognitive aids designed for use both within and outside anaesthesia. The only attributes that could not be scored were the topic-specific items where the investigators could not comment. It should be emphasised that it was not our intention to provide an evaluation of the clinical content or comment on best practice. The tool demonstrated areas of improvement for all the documents. Of note, we could find no explicit mention of design considerations in the shoulder dystocia or asthma guidelines. In the case of the asthma guidelines, the external specialist reviewers were explicitly asked 'to comment primarily on the comprehensiveness and accuracy of interpretation of the evidence base supporting the recommendations in the guideline' [25]. Conversely, the Resuscitation Council (UK) is explicit in its approach: 'As the guidelines will be used in emergencies, where efficient, timely action is critical, they should include clear, succinct recommendations with easily understood algorithms. Considerable care is taken to ensure that the guidelines are written plainly and unambiguously; this includes review by non-medical individuals before publication' [23].

We took a well-established tool designed to assess drills and checklists used on aircraft and developed and validated a unique tool for assessing the design of cognitive aids in medicine. We believe CMAT provides an objective and informative assessment of important human factor characteristics and will guide developers

in creating cognitive aids usable under the unique time pressures of medical emergencies.

Acknowledgements

We thank Dr Jane George, Dr Mary O'Connor, Dr Deborah Hinchliffe and Dr Kathryn Corrie for acting as independent assessors.

Competing interests

No external funding and no competing interests declared.

References

1. Lipitz-Snyderman A, Steinwachs D, Needham DM, Colantuoni E, Morlock LL, Pronovost PJ. Impact of a statewide intensive care unit quality improvement initiative on hospital mortality and length of stay: retrospective comparative analysis. *British Medical Journal* 2011; **342**: 219.
2. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine* 2009; **360**: 491–9.
3. de Vries EN, Prins HA, Crolla RM, et al. Effect of a comprehensive surgical safety system on patient outcomes. *New England Journal of Medicine* 2010; **363**: 1928–37.
4. Haugen AS, Softeland E, Almeland SK, et al. Effect of the World Health Organization Checklist on Patient Outcomes: a Stepped Wedge Cluster Randomized Controlled Trial. *Annals of Surgery* 2014 May 13; doi 10.1097/SLA.0000000000000716.
5. Harrison TK, Manser T, Howard SK, Gaba DM. Use of cognitive aids in a simulated anesthetic crisis. *Anesthesia and Analgesia* 2006; **103**: 551–6.
6. Ziewacz JE, Arriaga AF, Bader AM, et al. Crisis checklists for the operating room: development and pilot testing. *Journal of the American College of Surgeons* 2011; **213**: 212–7.e10.
7. Marshall S. The use of cognitive aids during emergencies in anesthesia: a review of the literature. *Anesthesia and Analgesia* 2013; **117**: 1162–71.
8. Degani A, Wiener EL. Cockpit checklists: concepts, design, and use. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 1993; **35**: 345–59.
9. Burian BK. Emergency and abnormal checklist design factors influencing flight crew response: a case study. *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics 2004* 2004; **1**: 6.
10. Civil Aviation Authority. Guidance on the design, presentation and use of emergency and abnormal checklists. www.caa.co.uk/docs/33/CAP676.PDF (accessed 01/06/2014).
11. Burian BK. Design guidance for emergency and abnormal checklists in aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2006; **50**: 106–10.
12. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS ONE* 2011; **6**: e20476.
13. Stanford Anesthesia Informatics and Media Lab. Cogaids. http://www.anesthesiaillustrated.org/cogaids/download-airway-emergency-aids-part-1 (accessed 01/06/2014).
14. Amathieu R, Combes X, Abdi W, et al. An algorithm for difficult airway management, modified for modern optical devices (Airtraq laryngoscope; LMA CTrach™): a 2-year prospective validation in patients for elective abdominal, gynecologic, and thyroid surgery. *Anesthesiology* 2011; **114**: 25–33.
15. Borshoff D. *The Anaesthetic Crisis Manual*. Cambridge: Cambridge University Press, 2011: 23–4.
16. Apfelbaum JL, Hagberg CA, Caplan RA, et al. Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. *Anesthesiology* 2013; **118**: 251–70.
17. Henderson J, Popat M, Latto I, et al. Difficult Airway Society guidelines for management of the unanticipated difficult intubation. *Anaesthesia* 2004; **59**: 675–94.
18. Project Check. http://www.projectcheck.org/crisis.html (accessed 01/06/2014).
19. Stephens CT, Kahntroff S, Dutton RP. The success of emergency endotracheal intubation in trauma patients: a 10-year experience at a major adult trauma referral center. *Anesthesia and Analgesia* 2009; **109**: 866–72.
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**: 297–334.
21. Duhachek A, Coughlan AT, Iacobucci D. Results on the standard error of the coefficient alpha index of reliability. *Marketing Science* 2005; **24**: 294–301.
22. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; **20**: 37–46.
23. Nolan JP. *Resuscitation Guidelines 2010*. London: Resuscitation Council (UK), 2005.
24. Royal College of Obstetricians and Gynaecologists. *Green-top Guideline No. 42*, 2nd edn. London: Royal College of Obstetricians and Gynaecologists, 2012.
25. British Thoracic Society. http://www.brit-thoracic.org.uk/document-library/clinical-information/asthma/btssign-guideline-on-the-management-of-asthma/ (accessed 01/06/2014).
26. Jenkins B. Cognitive aids: time for a change? *Anaesthesia* 2014; **69**: 660–4.
27. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical Education* 2003; **37**: 830–7.
28. Downing SM. Reliability: on the reproducibility of assessment data. *Medical Education* 2004; **38**: 1006–12.
29. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2011; **2**: 53–5.
30. Tighe J, McManus I, Dewhurst NG, Chis L, Mucklow J. The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP (UK) examinations. *BMC Medical Education* 2010; **10**: 40.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
32. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy* 2005; **85**: 257–68.
33. Malone DC, Abarca J, Hansten PD, et al. Identification of serious drug-drug interactions: results of the partnership to prevent drug-drug interactions. *Journal of the American Pharmacists Association* 2004; **44**: 142–51.
34. Strasser S, London L, Kortenbout E. Developing a competence framework and evaluation tool for primary care nursing in South Africa. *Education for Health* 2005; **18**: 133–44.
35. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Medical Research Methodology* 2005; **5**: 37.

Appendix

The Cognitive aids in Medicine Assessment Tool (CMAT)

Title	Attribute	Description	Score
Domain 1 physical characteristics			
Document size			
1.1	Is the size of the document appropriate to the space available?	The cognitive aid must be visible in an accessible location and of an appropriate size	
Tabs and dividers			
1.2	Are any tabs that are used clearly identified?	Tabs are physical attributes that allow quick identification and access at the beginning of a section/chapter. Tabs should be clearly identified in order to avoid a delay in locating the correct drills. Tab numbering should be consistent throughout the document	
Font type			
1.3	Does the font type used provide clear differentiation between characters?	Use of sans serif fonts (without tails) such as Helvetica, Gill Medium or Arial fonts are recommended as these are easier to read	
Print size			
1.5	Are the action points legible at arms' length?	The text must be legible under all lighting conditions at arm's length (approximately 600 mm)	
Margins			
1.6	Can you use your thumb as a cursor to keep track of progress through the cognitive aid?	It should be possible to hold the list using the thumb as a cursor without obscuring the text. A 19-mm margin is recommended	
1.7	Are all steps aligned to left?	If the steps run horizontally or diagonally away from the left margin it will not be possible to use thumb as a cursor.	
Contrast and colour			
1.8	Has black text on a white or yellow background been used? 'Alert cues' may be coloured	Recommend using white or yellow background. If other colours are used check legibility under low ambient lighting	
1.9	Where colour shading has been used to discriminate actions or notes, is there sufficient contrast between the text and background?	Colour shading provides a good method of discrimination but must be used with care. The use of pastel colours (low saturation) for shading is recommended.	
Numbering			
1.10	Are page numbers clearly identified?	Lack of page numbers can make the cognitive aid unusable. Recommend putting the number at the bottom or top of the page with a large font size. If the document has only a single page, then a statement similar to '1 of 1' is recommended	
1.11	Are *actions* consecutively numbered?	Numbering actions assists in place keeping. Multiple actions need to be numbered	

Appendix (continued)

Title	Attribute	Description	Score
Domain 2 content			
Structure			
2.1	Has the number of action items been minimised to take account of time available to complete the cognitive aid?	Actions that only offer a small chance of success may not be essential when time is critical	
Title			
2.2	Does the cognitive aid have a title?	The title must stand out from the action items and notes. Tools such as boxing or using bold text are recommended	
2.3	Does the title fully reflect the failure condition?	The title should be succinct and unambiguous in order to avoid the use of the incorrect cognitive aid. Also the title should be appropriately summarised without compromising the understanding of the failure condition	
Failure condition			
2.4	Does the cognitive aid contain a description of the failure condition(s)?	The description may be separate from the title or included within the title. A description of the failure conditions is recommended as this provides a useful confirmation that the correct cognitive aid has been selected	
Objective			
2.5	Does the cognitive aid contain an objective?	An objective statement is recommended where appropriate, as it is a useful confirmation that the correct cognitive aid has been selected and what the expected outcome is	
Memory items			
2.6	Are any memory items used listed at the beginning of the cognitive aid?	Memory items are recall items i.e. actions that must be performed very quickly in response to a failure when there is no time to refer to a document. Memory items should be carried out first	
2.7	Are any memory items clearly distinguished from the other action items?	It is recommended that the memory items be distinguished in some fashion – boxing, shading, line marking, numbering (M1, M2), etc	
2.8	If used, are there six or fewer memory items?	It is recommended that the memory items should be kept to a minimum – preferably four or fewer. Recall can be impaired under stressful situations	
Cautionary notes			
2.9	Are any cautionary notes clearly discriminated?	Cautionary notes highlight resultant performance constraints and should be differentiated from ordinary explanatory notes. It is recommended that appropriate colour shading highlights cautionary notes. Ideally, they should be accompanied by the word 'caution'	
2.10	Are any cautionary notes printed above the action item to which they relate?	It is essential that the person using the cognitive aid is aware of the implications of any action item before they carry it out. Recommend moving the cautionary note to precede the action to which it relates	

Appendix (continued)

Title	Attribute	Description	Score
Action items			
2.11	Are any action items used distinguishable from the text?	It is important to identify the items in the list where action is needed. Recommend that they are distinguished from other items in the document (e.g. by text font size, font type or bold font).	
2.12	Are the 'read' and 'do' items clearly linked?	This attribute applies to the 'challenge and verification' method of conducting a checklist. The use of dots or dashes to link challenge and response items is recommended to avoid the possibility of associating the wrong challenge and response	
2.13	Are any critical items discriminated?	Critical items are ones that, if not carried out after a certain event has occurred, may lead to a hazardous situation. It is recommended that critical items should be presented in a way that discriminates them from action items. Critical actions usually require positive confirmation from other team members	
2.14	Where appropriate, does the procedure explicitly state who is responsible for specific actions?	The instructions should indicate who is responsible for carrying out specific actions	
Explanatory notes			
2.15	Are any explanatory notes clearly distinguished from action items?	The notes should not clutter the action items. It is recommended that they are visually distinguishable	
2.16	Are those notes linked to the action item to which they relate?	It is essential that the notes either precede or follow the action item. It is recommended that notes are close to and consistently placed relative to the action items to which they refer. Inconsistent placing may be confusing to the user	
Decision items			
2.17	Are conditional steps clearly laid out?	Complicated conditional statements, particularly with action items embedded within them, may lead to error. It is recommended that complex conditional steps are minimised in emergency situations to avoid cognitive overload.	
Review of system status			
2.18	Is a review of the clinical situation provided?	An explicit requirement to reassess the situation following an intervention or at each new stage of an algorithm helps to keep people engaged, reduce task fixation and distribute awareness	
Deferred items			
2.19	Is the presence or absence of deferred items clearly identified and necessary actions described?	Are deferred items, such as need for specialist referral, clearly identified and described? For example, by using a 'who/when/where' format	
3 *Layout and format*			
Cognitive aids per page			
3.1	If the cognitive aid runs onto a second page, is it split at a logical place?	Cognitive aids should be split into logical sections and the logical sections should not be split at a page break as this reduces continuity	

Appendix (continued)

Title	Attribute	Description	Score
Start and finish			
3.2	Does the cognitive aid have a clearly defined start?	The cognitive aid will be unusable if it is not clear where it starts. It must have a clearly defined start	
3.3	Does the cognitive aid have a defined end?	The end of the document must be indicated with an 'end of xxx' indication or graphical equivalent	
3.4	Are the 'end of xxx' indications provided in every place where the cognitive aid can be completed?	The end of cognitive aid statement must be included at all places when it is complete	
Continuation pages			
3.5	Is it clear when the cognitive aid continues on to another page?	The cognitive aid may not be completed if it is not clear that it continues on to another page. It is recommended that a clear indication be provided at the bottom of the page and top of the continuing page	
Order			
3.6	Does the order of the action items ensure return to a safe state at the earliest opportunity?	The design of the cognitive aid must ensure that priority items, i.e. those that will deal with the fault in the most time efficient way, are in the appropriate order	
Cross-referencing			
3.7	Is cross-referencing minimised?	Internal cross-referencing can lead to error if it is not clear to which step it refers. It is recommended that the use of cross-referencing is minimised and that steps are numbered when cross-referencing is used	
3.8	Where there is cross-referencing to other material is it appropriately signposted?	If cross-referencing is crucial then it is recommended that a place-keeper (in the form of an asterix or other symbol) be used to aid return to the correct place in the cognitive aid. It is also recommended that the document and page number are clearly referenced	
Figures and tables			
3.9	Are any figures or tables clearly linked to the cognitive aid with which they are associated?	Errors will occur if the wrong figures or tables are referred to. It is recommended that the figures and tables should be clearly labelled to allow correct referencing	
3.10	Are the figures legible and usable?	Data contained in graphs or tables will not be usable if the presentation is too small. They should be readable at an arm's length	
Abbreviations and consistency			
3.11	Do all captions and labels used in the cognitive aid correspond exactly to the words used in the clinical environment?	It is essential that exact correspondence is achieved and any differences must be corrected	
3.12	Does the cognitive aid include a statement of currency (i.e. is it in date)	If not included this could result in use of the wrong cognitive aid or unavailable equipment being sought	
3.13	Can the cognitive aid be made site-specific?	Consideration should be made to providing customisable sections so that the document can be made site-specific	
Total			