Validation for a scoring system of the ALS Cardiac Arrest Simulation Test (CASTest)

Fiona Napier[1,2], Robin P Davies[2,3], Catherine Baldock[5], Harry Stevens[6], Andrew S Lockey[3,6], Ian Bullock[3,7], Gavin D Perkins[1,2,3]

[1] Warwick Medical School, Gibbett Hill, Warwick, CV4 7AL

[2] Heart of England Foundation Trust, Bordesley Green East, Birmingham, B9 5SS

[3] Resuscitation Council (UK), Tavistock House North, Tavistock Square, London, WC1H 9HR

[4] Prince Charles Hospital, Merthyr Tydfil, CF47 9DT

[5] University Hospital Coventry and Warwickshire, Clifford Bridge Road Coventry, CV2 2DX

[6] Calderdale & Huddersfield NHS Foundation Trust, Halifax, HX3 0PW

[7] Royal College of Nursing Institute, Whichford House, Oxford OX4 2JY

Keywords: Advanced Life Support; Cardiopulmonary Resuscitation; Training; Assessment; Validity

Corresponding Author:

Dr. Gavin D Perkins

Email: g.d.perkins@warwick.ac.uk

Fax: +44 (0) 247657 4657

Abstract: 249 words

Manuscript: 2472 words

Abstract

Aim: The Cardiac Arrest Simulation Test (CASTest) assesses resuscitation knowledge and skills during a simulated cardiac arrest. The aim of this study is to validate an alternative scoring system for measuring individual candidate performance during research involving the CASTest.

Methods: The performance of 537 participants was measured using the new scoring system. In addition, assessors assigned a global pass-fail decision. Differences in scores were compared between to global pass/ fail decisions, professional groups and those nominated to become instructors. Correlations between CASTest domain scores and overall score, multiple choice scores and other practical tests (airway test and initial assessment and resuscitation test) scores were measured. This provided opportunity to cross reference achievement in other areas of course assessment with this alternative scoring system.

Results: 413 (76.9%) passed the CASTest and 124 (23.1%) failed. The total performance score was significantly higher in those that passed than in those that failed (median 77 vs 62.5, P<0.0001). There were no differences between professions. Senior staff performed slightly better than junior staff (median 74 and 72 respectively, P=0.01). Excellent participants (identified as having instructor potential) scored significantly higher than the other participants (median 94 and 72 respectively, P <0.0001). A strong correlation was demonstrated between domains in the CASTest (rho 0.72-0.82, P<0.01). Other assessment outcomes for the ALS course correlated poorly with CASTest scores (rho = 0.27-0.37, P<0.01).

Conclusion: This new simple scoring system can be used to better characterise performance on the ALS course CASTest than the current binary pass-fail outcome.

1. Introduction

A variety of healthcare professionals need to demonstrate competency in all aspects of Advanced Life Support (ALS).[1, 2] It is vital that certification of such skills is based on robust testing, giving ALS providers credibility and enabling them to promote patient safety. As with any education test, evidence of validity is key if the results are to be interpreted meaningfully.[3] Assessments are not valid or invalid in themselves but rather a certain interpretation is more or less valid for a certain population at a certain point in time.[4] Evidence should be collected from multiple sources and analysed to create arguments for and against a specific interpretation of test results.

The ALS course teaches both theoretical and practical aspects of resuscitation. Participants complete a pre-course Multiple Choice Question (MCQ) test based on pre-course preparation. During the course participants are assessed on practical skills involving airway management and the initial approach to critically ill patients. Summative assessment at the end of the course is a combination of an MCQ paper (pass mark 75%) and a cardiac arrest simulation test (CASTest).[1]

The CASTest uses a simulated cardiac arrest to test the application of resuscitation knowledge, and skills and is focussed on team leadership and decision making.[5] The score sheet contains performance criteria that participants are expected to demonstrate in order to pass the test. Performance criteria are classified as bold (i.e. essential) and non-bold (desirable). Based on this structured assessment of performance the assessment outcome is reported as a single binary pass-fail mark. We have developed a four point scoring system which can be applied to each of the 24 performance criteria in order to characterise the quality of candidate performance in more detail. The scoring system does not give differential score for bold and non-bold treatment points.

The aim of this study is to determine the measurement properties of the new tool by considering the tool's internal structure and relationship with other variables (i.e. theoretical (post course MCQ paper) and practical skill assessments) and thus report it's utility to better characterise performance in CASTest than a simple pass / fail result.

2. Methods

2.1 Participants

The study was approved by South Birmingham Research Ethics Committee. Participants enrolled in an evaluation of pre-course computer simulation material from ALS courses at 11 UK centres were eligible for inclusion in the study. Participants provided written informed consent. The study was conducted between March and December 2007.

2.2 Assessment criteria

Participants underwent assessment of the following aspects: pre/post course MCQ paper, skills assessments (airway, initial assessment and resuscitation), and a CASTest (CASTest- scenario 1). This study evaluated data from the first attempt at CASTest only.

The CASTest assesses performance against 24 performance criteria. These cover 4 domains: initial assessment and resuscitation (5 criteria) and cardiac arrest management - PEA (7 criteria); VF (11 criteria) and post resuscitation care (1

criterion). A new four point scoring scale was assigned to each performance criteria to replace the previous “achieved” or “not-achieved” criteria. Each assessment was undertaken by two Resuscitation Council (UK) trained assessors. These assessors had been involved in training prior to assessment. Assessors evaluated performance individually and agreed a joint score by consensus between the pair. If they failed to agree on a score the Course Director acted as the final arbiter. Assessors were provided with written guidance as follows:

4 - The highest score is awarded to excellent participants who made correct decisions promptly and with confidence, demonstrating expert performance and instructor potential.

3- Acceptable performance should score 3. This is the usual level of competence attained by an ALS provider. They were able to make correct decisions, but may have some hesitation or lack of confidence.

2 - Borderline performance. Minor errors in decision-making, hesitant, lacked confidence and required prompting or failed to perform a skill but recognised errors on subsequent questioning.

1 – Unacceptable. Participants who make incorrect decisions or gave inappropriate treatments. Their actions may have caused harm in a real life situation.

At the end of the CASTest instructors provided a global assessment as to whether performance was acceptable or not (pass / fail) and overall grade of performance (using the 4 point assessment scale defined above). The mark sheet can be viewed in the electronic supplemental material.

The pre and post course multiple choice questions are two different 120 item true false multiple choice questions. Individual questions are grouped into blocks of four with a common stem. Evaluation from over 5000 multiple choice questions for each paper have demonstrated excellent agreement (personal communication Carl Gwinnutt).

Airway management and initial assessment and resuscitation assessments are skill based assessments. These skills are assessed by outcome based assessment i.e. participants are allowed as many assessment attempts during the course as required to allow them to achieve the necessary standard. The airway assessment tests basic airway care, the use of simple airway adjuncts and laryngeal mask insertion. The initial assessment and resuscitation station assesses the management of a critically ill patient at risk of cardiac arrest, delivery of CPR and safe defibrillation. Performance in these assessments were also rated using the same 4 point scale described for the CASTest, to provide cross referencing against candidate assessment profiles.

At the end of each course, the assessors met as a group to identify participants with outstanding performance who may be considered to train as future instructors (instructor potential). Participants performance over the course were considered using a structured scoring sheet (MCQ score, communication, enthusiasm, ability to critique (self and other participants), interactive, supportive, team member and credibility). The assessors did not have access to the CAStest total performance score during these deliberations.

2.3 Data analysis

SPSS statistical package version 15.0 was used for analysis of the data. Data were assessed for normality and found to have a non-parametric distribution. Comparisons between the scores of participants that passed and failed were analysed using the Mann Whitney U Test. In addition scores of senior and junior staff were compared, as well as those with and without instructor potential. Differences between scores and overall grading (scores 1-4) were measured by Friedman's test[6].

Scores for the three main domains (initial approach; PEA, VF) within the CASTest were examined for associations as it was hypothesized that participants performing well in one domain would also perform well in others (internal structure). For example those scoring highly in ventricular fibrillation management would also score highly in pulseless electrical activity management. Such correlations were measured using Spearman's rho, aiming to provide evidence of internal structure of the test.

The relationship between total CASTest score and other course outcomes (pre-course MCQ, end of course MCQ and practical skill station scores (airway and initial assessment / resuscitation)) was measured to gain evidence relationship to other variables.

3. Results

A total of 537 participants were assessed: 346 doctors, 97 nurses, 7 operating department practitioners, 8 others and 79 unknown. There were more than double the number of junior staff compared to senior staff (267 vs 114), whilst only 27 were students and 129 unknown. Data for pre-course MCQ were available from 429 (79.9%) of participants.

Evidence of internal structure

Overall 413 (76.9%) participants were awarded a pass by the assessors and 124 (23.1%) were assigned a fail according to the standard assessment criteria.. The total performance score was significantly higher in those that passed the test than in those that failed (median 77 (IQR 72-92) vs median 62 (IQR 56-68) P<0.0001) (figure 1). Scores ranged from 35 to 96 (the maximum possible). The CASTest score related well to the global assessment of performance (figure 2).

There were significant positive associations between scores for all CASTest domains (Figure 3). For example, there was a strong correlation between the score for initial approach to the critically ill patient and scores for ventricular fibrillation management. Correlation coefficients ranged between 0.72-0.82 (P<0.01) for all comparisons.

Evidence of relationships with other variables

There was no significant difference in CASTest scores between doctors and nurses (median 73 vs 72, P = 0.816). Numbers of other health professionals were not adequate to meaningfully interpret their scores in comparison. Senior staff performed slightly better than junior staff (median score 74(IQR 69-93) vs 72 (IQR 69-84); P = 0.01). 45 potential ALS instructors were identified at the end of the courses. They achieved significantly higher scores than the other participants (median: 94 vs 72, P<0.0001).

There were significant, albeit weak correlations between CASTest score and the other assessment outcomes. Post-course multiple choice examination scores only correlated slightly better with CASTest scores than pre-course multiple choice scores (Spearman's rho = 0.336 and 0.269 respectively, P<0.01 for both). Airway management and the initial assessment and resuscitation of the patient also had poor associations with CASTest score (Spearman's rho = 0.325 and 0.367 respectively, P<0.01 for both).

4. Discussion

This study presents evidence supporting the validity of a scoring system which can be used to measure overall performance during advanced life support cardiac arrest simulation testing. This system will allow greater precision in classifying performance than the existing binary pass or fail outcome measure. Using a large cohort of multi-professional health care providers the participants awarded a pass mark by the standard assessment tool obtained significantly higher performance scores than those that failed. There were differences in total score between global assessments of performance measured by a 4 point scale. There was good internal consistency between the treatment domains of initial assessment, management of PEA and management of VF. Although significant associations were present between CASTest score and other assessments such as the multiple choice question papers, airway and initial assessment and resuscitation skill stations, these were less strong.

It is important that assessments used in today's healthcare setting are fit for purpose. There are a number of different measures that can be used to judge the performance of a test. These include the reliability, validity, feasibility, cost effectiveness, specificity and fidelity.[7] Downing describes all validity as construct validity, for which

there are five distinct sources of evidence:, internal structure (statistical or psychometric properties of the score), relationship to other variables (the ability to discriminate between levels of training) content (whether the test assesses learning objectives), response process (data integrity) and consequences.[4] The study sought evidence of internal structure and relationships with other variables.

The study identified evidence of internal structure by showing highly significant differences in scores between participants who passed and failed the CASTest and between global ratings of performance during performance. There were also strong correlations between scores within different CASTest domains. Evidence of relationships with other variables was derived from the findings that participants deemed to have instructor potential did significantly better than other participants. There were also statistically significant differences between junior and senior staff performance although the clinical significance is likely to be minimal.

The CASTest score did not correlate as strongly with performance on the pre and post course multiple choice tests. A number studies show poor agreement between tests measuring theoretical knowledge and demonstration of practical resuscitation tests.[8-10] This is likely to reflect the different cognitive processes involved in knowledge and skill tests. However one might expect better correlation between CASTest and other practical tests such as airway management and initial assessment and resuscitation. One explanation could be the complexity of the skills being tested. Airway and initial assessment and resuscitation skills are more basic task orientated assessments. CASTest demands the integration of theoretical knowledge, patient assessment, reasoning and clinical skills which require a higher level of cognitive functioning. Alternatively this poor association may be explained by differences in the assessment process. Participant scores for the CASTest were derived from a single assessment of performance whilst assessment of airway and

initial assessment and resuscitation skills are assessed over a period of time. During this, participants are not penalised for errors provided they demonstrate the skill eventually. The absence of strong associations provides some reassurance that results were not contaminated by a halo effect (assessors marking candidates based on previous encounters of performance during the course rather than actual performance during the CASTest)

The CASTest tests several of the key learning outcomes for the ALS course. These include the ability to recognise and intervene in the management of a simulated patient at risk of cardiac arrest; lead a team in the resuscitation of a simulated patient in cardiac arrest; demonstrate knowledge and application of current resuscitation guidelines, demonstrate an understanding of the importance of post-resuscitation care and stabilisation following a return of spontaneous circulation. This provides evidence of content validity.

The demonstration of construct validity in the present study are supported by similar findings from Ringsted *et al.*[11] Using the same CASTest clinical scenario but with a 5 point rating scale, the authors were able to differentiate between novice and advanced learner performance. Further evidence of the construct validity of the test comes from an evaluation of CASTest outcomes in over 2000 ALS providers. This study demonstrated similar pass rates for the four different CASTest assessments whilst was able to detect differences in outcomes according to professional background.[12]

Reliability was not specifically assessed in this study. One of the first studies to examine reliability during CASTest found evidence of poor inter-observer agreement of videotaped CASTest scenarios.[13] Agreement subsequently improved following the

introduction of standardised performance criteria and paired assessor marking.[14] The most recent assessment of reliability during CASTest assessment comes from Ringsted *et al.*[11] This study found high levels of intra and inter observer agreement for performance criteria (intraclass correlation coefficients ranging 0.84-0.97) and moderate levels of inter-observer agreement for the overall pass/fail decision (average kappa 0.72).

The study has a number of limitations. These include sampling method, which was confined to 11 centres in the UK. However pass rates were comparable to a previous evaluations[12] and data on file at the Resuscitation Council (UK). The ALS course is used throughout Europe and whilst we have not reason to suppose the scoring system would perform differently outside the UK, this remains a possibility. The evidence of relationship with other variables was limited to comparisons between CASTest score and performance in MCQ's, other practical skill stations and faculty assessment of instructor potential. Additional evidence of external validity could be sought by linking CASTest scores to performance in other simulated emergencies, peer assessments of performance and real life resuscitation attempts. A strength and limitation of the score is that it measures overall performance during the CASTest. The score will not be particularly sensitive at identifying an otherwise high performing candidate that commits a single critical error. However as the purpose of the tool is to measure overall performance this is of a lesser significance than if it was being proposed as a tool to determine the overall assessment outcome. Finally the evaluation was limited to CASTest scenario 1. Although the other CASTest scenarios use the same performance criteria, the performance of the tool with these CASTests has not been demonstrated.

The strength of the CASTest score is the ability to rate candidate's performance with more precision than the existing binary pass fail outcome. This has the advantage that it will allow smaller sample sizes to be used in future research studies evaluating performance during CASTest. The benefits during routine ALS courses need to be determined. Whether introducing the CASTest score with a specific cut-off score to differentiate between acceptable and un-acceptable will improve reliability requires further investigation.

Conclusions

Evidence supporting the construct validity of the CASTest scoring system is presented. This simple scoring system better characterises performance in the ALS course CASTest than the current binary pass-fail outcome.

Acknowledgments

We would like to thank the Resuscitation Research Group for assistance with the project: Nicole Gomez-Davis, Marie Fletcher, Teresa Melody, Fang Gao and Dawn Hill.

GDP holds a DH (NIHR) Clinician Scientist and is supported by the Intensive Care Foundation. The project was funded by the Resuscitation Council (UK) and Laerdal Foundation for Acute Medicine. The study design, analysis and decision to publish was the responsibility of the authors.

Collaborators

Julie Lawrence, Good Hope Hospital, Birmingham

Ben King, Gloucester Hospital

John Heard, City Hospital, Birmingham

Clare Green, Warwick Nuffield Hospital

Legends to figures

1) Total performance score is significantly higher in participants that pass than those that fail the CASTest assessment ($P<0.001$).

2) CASTest score increases sequentially with increases in global assessment of performance during the CASTest assessment

3) There are strong correlations between scores for CASTest domains (initial approach, pulseless electrical activity and ventricular fibrillation management). Spearman's rho = 0.72-0.82, $P<0.01$.

4) CASTest score correlates poorly with the other assessment outcomes (airway management, initial assessment and resuscitation, pre- and post-course MCQ scores).

References

1. Perkins G, Lockey A. The advanced life support provider course. BMJ 2002;325:S81.
2. Baskett PJ, Nolan JP, Handley A, Soar J, Biarent D, Richmond S. European Resuscitation Council guidelines for resuscitation 2005. Section 9. Principles of training in resuscitation. Resuscitation 2005;67 Suppl 1:S181-9.
3. Norcini JJ. Setting standards on educational tests. Medical Education 2003;37:464-9.
4. Downing SM. Validity: on the meaningful interpretation of assessment data. Medical Education 2003;37:830-7.
5. Perkins GD. Simulation in resuscitation training. Resuscitation 2007; In Press
6. Perkins GD. Mann Whitney U test. In: Gao SF, Smith JE, eds. Key Topics In Medical Research. London: Bios; 2002:128-32.
7. Resuscitation C, Bullock I, Colquhoun M, Coleman A. The assessment of resuscitation teaching. In: The Generic Instructor Course. London: Resuscitation Council (UK); 2004:34-41.
8. Smith KK, Gilcreast D, Pierce K. Evaluation of staff's retention of ACLS and BLS skills. Resuscitation 2008;78:59-65.
9. Lockyer J, Singhal N, Fidler H, Weiner G, Aziz K, Curran V. The development and testing of a performance checklist to assess neonatal resuscitation megacode skill. Pediatrics 2006;118:e1739-44.
10. Nadel FM, Lavelle JM, Fein JA, Giardino AP, Decker JM, Durbin DR. Assessing pediatric senior residents' training in resuscitation: fund of knowledge, technical skills, and perception of confidence. Pediatric Emergency Care 2000;16:73-6.
11. Ringsted C, Lippert F, Hesselfeldt R, et al. Assessment of Advanced Life Support competence when combining different test methods--reliability and validity. Resuscitation 2007;75:153-60.
12. Perkins GD, Davies RP, Stallard N, Bullock I, Stevens H, Lockey A. Advanced life support cardiac arrest scenario test evaluation. Resuscitation 2007;75:484-90.
13. Perkins GD, Hulme J, Tweed MJ. Variability in the assessment of advanced life support skills. Resuscitation 2001;50:281-6.
14. Perkins GD, Tweed MJ. Evaluation assessment of advanced life support skills: examiner consistency. Med Educ 2001;35:S60.