

Reliability in Assessing Basic Life Support Skills in infants by Utilizing Video Recordings

Salman Saleem*
Rizwan Hameed Malik**
Shamsa Rizwan***
Assad Hafeez****
Farrukh Qureshi*****

*Consultant physician and APLS instructor
CMH Gujranwala

**Assistant Professor of Orthopaedics
Pakistan Institute of Medical Sciences
Islamabad

***Professor of Gyn/Obs, Riphah University
Rawalpindi

****Consultant Pediatrician & Director
Health Services Academy, Islamabad

*****Program Officer
WHO EMRO

Background: In the assessment of Basic Life support Skills taught to the health workers, the use of video recordings can be utilized in the virtual learning environment if proved to be a reliable method.

Aims: This study analyses whether video recording is a reliable tool in basic life support skills assessment.

Materials and Methods: The health workers trained in basic life support skills were assessed while performing the skill on an infant manikin. The skill was recorded by a camera and was directly observed by an instructor who marked the performance on a BLS skill assessment instrument in the form of a numeric score as well as an overall pass fail score. The recordings were coded and two ALSG certified instructors viewed the video recordings independently, rating every BLS performance using the same assessment sheet that was used for on-site assessment.

Results: There was good correlation between the two video assessments (ICC 0.957) as well as among the observed assessment and the two video assessments (ICC 0.973 with 95% CI; 0.959 – 0.982 and 0.957 with 95% CI; 0.934 – 0.971).

Conclusion: Video recording of basic life support skills is a reliable assessment tool.

Address for Correspondence:

Dr Rizwan Hameed Malik
House#9/B, Street 31, F-8/1,
Islamabad 44000, Pakistan
E mail: drrizwan@nayatel.pk

Introduction

Basic Life Support Skills (BLS) is an essential skill for medical practitioners and the assessment of the health worker's competence cannot be underestimated. There can be various methods of assessment of a skill, hands on performance on a manikin is one of the commonly used methods. The role of video recordings in medical education is already recognized in teaching, training and to some extent in assessment. In our study we propose to test the reliability of video recordings in the assessment of BLS skills.

Pakistan is a developing country and the Emergency Maternal and Child Health (ESS-EMCH) course is a life-support training designed specifically for health workers by Advanced Life Support Group (ALSG), UK.^{1,2} The qualified instructors are expected to conform to uniform teaching and assessment protocols as laid down by ALSG-UK. The local instructors have been conducting the ESS-EMCH courses in Pakistan since January 2005. One of the essential components of these courses is to teach Basic Life Support (BLS) skills according to internationally accredited BLS protocol.

This study compared the on-spot live assessment of a set of randomly selected BLS

performances to the re-assessment made on the video records of the same performances from various courses. This study was conducted as part of a comprehensive evaluation of ESS-EMCH courses conducted in Pakistan during 2006.

Materials and Methods

The ESS-EMCH courses were attended by a mixed cohort of general practitioners, paediatricians, obstetricians and nurses belonging to various governmental health facilities of Pakistan. An informed consent was obtained from all the participants prior to video recording. The first day of the course is dedicated to teaching BLS skills which are taught according to the ALSG protocols used in the Advanced Paediatric Life Support Course (APLS).^{3,4} The assessment of BLS skills is done at the end of the day by the instructors.

To determine the inter-rater agreement, it was presumed that a standardized training methodology should not allow a difference of opinion of more than 30% (a correlation > 0.3) amongst the raters. The number of cases (performances) required to detect any correlation difference of more than 0.3 against a null

hypothesis value of 0 (no difference of opinion) was calculated to be 85 with 80% power at $\alpha = 0.05$.⁵

The participants were asked to perform BLS skills on an infant manikin. The skill was observed by an instructor who marked the performance on a BLS skill assessment instrument (Figure 1). The BLS skill assessment instrument used in our study was adapted from the standard checklist used in the ALSG courses. The instrument assessed the candidates for performing nine essential steps of resuscitation in the desired sequence. The final verdict of being 'successful resuscitation' was not only based on accomplishing every task in desired sequence referred to as 'structured approach', but also as to whether or not the BLS skills were performed correctly.

Figure 1: BLS Skill Assessment Instrument used in ESS-EMCH Course.

ESS-EMCH COURSE BASIC LIFE SUPPORT TEST: INFANT		
MARKING SHEET		
Candidate Name: _____		Candidate No: _____
	Task	☒
1	SAFE Approach *	☐
2	Are you alright?	☐
3	Airway opening maneuvers	☐
4	Look, listen, feel	☐
5	5 initial rescue breaths	☐
6	Pulse check for 5-10 seconds	☐
	Chest compressions	
7	Finger/hand position	☐
8	Technique	☐
9	Ventilation	☐
10	Ratio 15:2 rate of 100 per minute Including "missed" beats.	☐
11	Call emergency services**	☐
PASS ☐		
FAIL ☐		
Signed: _____		
*Excluded from analysis: not shown in video		
**Excluded from analysis; required verbal response from examinee		

The assessment was marked in the form of a numeric score by awarding one mark for every step accomplished. The performance was also categorized as either 'successful' or 'unsuccessful' at the completion of the exercise. The performances of course participants were also recorded on a video camera mounted on the side of the BLS skill station. Care was taken not to show the faces of the examinees, with the camera focused on the manikin, to protect confidentiality and to avoid potential assessment bias.

All the video recordings of BLS performances before and after the training sessions were collected from the four courses conducted from March to December 2006. The recordings were appropriately coded and grouped in the form of one set comprising clips in a random order. The videos with inappropriate focus, or recording errors were discarded from final cohort.

The purpose of the study and the recording procedure was explained to the two ALSG certified instructors who did not teach on any of the included courses, and consented to participate in the study. Three samples of video recordings were shown to the instructors, to address any queries, and to validate that they understood the video assessment procedure. They were then asked to independently rate every BLS performance using the same assessment sheet that was used for on-site assessment.

The three evaluations; one on spot (OS; by various instructors), and the two video re-assessments (VA1 & VA2) were subjected to analysis to assess inter-rater variability.

Data Analysis Procedure: The inter-rater agreement was assessed on numerical scores obtained on the BLS skill assessment instrument as well as on the categorical rating of pass or fail.

The analyses were carried out in three steps. In the first step, the two video assessments were analyzed to ascertain the degree of conformity, and to validate video recordings as a tool to observe participants' performance. Each of the two video assessments was then compared individually with observed assessment in the second step. The three paired comparisons were analyzed in the third step to obtain a more plausible evidence of the level of conformity amongst the raters.

Data Analysis Technique: The concordance on cumulative numerical score (out of 9 maximum) was assessed by measuring Intra-class Correlation Coefficient (ICC) between the ratings. The phenomena of ICC as a method to rate agreement has been explained earlier by Shrout and Fleiss⁶, and later it's various indices and dimensions have been described by Robert Yafee⁷ and Portney et al.⁸ Our analyses used multi-rater ICC indices whereby judges were deemed as a random subset of multiple judges, and candidates as a random subset of group of candidates (Two-way

random effects model).⁷ Single measure ratings were used to calculate ICC and it was based on measuring the absolute agreement amongst various ratings. However ICC between the two video-taped ratings took into account the individual assessments by every rater, whereby the same raters had rated all the cases (Two-way mixed effect model). The ICC is a measure of the combined effect of within and between group's variance components (whereby the instructors and trainees were the two groups in our study). The ICC index has a maximum value of 1, with higher values representing more reliable data, or data in which the majority of variance is between groups or companies rather than within groups.⁹ The coefficient represents concordance, where 1 is perfect agreement (100%) and 0 is no agreement at all.

The inter-rater agreement on categorical marking the nine individual tasks of the BLS performance correct or wrong, and for the performance being 'successful' or not have been calculated by measuring Cohen's Kappa Correlation Coefficient (κ statistic).¹⁰ Kappa measures a concordance between multiple raters on a nominal scale across cases. It estimates the proportion of observed agreement among raters after removing the proportion of agreement which would occur by chance.¹¹ Cohens Kappa is usually calculated to measure agreement between two ratings. We used multirater κ statistic measurement expounded by Seigal and Castellan¹² for our raters' cohort. The upper limit of κ statistic is +1.00 and occurs when there is total agreement. A κ statistic of 0.00 indicates agreement at a chance level. Negative kappa values represent agreement which is less than chance (lower limit is -1.00). However, different assessment scales have been proposed to qualify the extent of agreement.^{13,14,15} We used the Krippendorff's scale to arrive at the conclusions along with 95% CI values. This scale disregards any rating with $K < 0.67$, allows tentative (adequate) conclusions when $0.67 < K < 0.8$ K, and definite (plausible) conclusions when $K > 0.8$ (Table 1).¹⁵

Results

Agreement between the two video ratings (VA1 & VA2): The ICC between the two video assessments was 0.957 (95% CI; 0.933 – 0.972)(Table II). The concordance on grading the performance as 'successful' resuscitation or not, was observed to be a Kappa value of 0.84 ($p = 0.00$). The κ -statistic for categorical rating of individual tasks ranged from 0.72 for ventilation breaths to 0.95 for CPR Ratio, with a median value of 0.88.

Comparison of the two video assessments with Observed assessment: The ICC between OS -VA1 was recorded at 0.973 (95% CI; 0.959 – 0.982), whereas it was 0.957(95% CI; 0.934 – 0.971) when OS-VA2 were analyzed. (Table 2)

The concordance on grading the performance was observed to a κ -statistic of 0.70 ($p = 0.00$) for OS-VA1 and 0.72 for OS-VA2 ($p = 0.00$).

The median value κ -statistics for the nine individual tasks amongst the two paired comparisons ranged from 0.73-0.95 for OS-VA1, and 0.66-0.93 for OS-VA2, with 'look, listen, feel' showing best concordance whereas 'airway opening' observation found to be least concordant. Figure II shows a comparison of κ statistics for the nine individual tasks amongst these three paired analyses.

Discussion

Video recordings have been used in medical education for various purposes like quality analysis¹⁶, to enhance physical examination skills¹⁷, for assessing competency in the clinical skills laboratory¹⁸, to test video-based communication skills¹⁹, for distance education courses²⁰, Objective Structured Video Examinations (OSVEs) for geriatrics²¹ and for the assessment of medical communication skills²². When used for assessment, videotape viewing significantly affected resident self-scores in professionalism,

Table I: Various Kappa Interpretation Scales

Interpretation of reliability					
Landis and Koch ¹³		Shrout ¹⁴		Krippendorff ¹⁵	
Value of Kappa	Strength of agreement	Value of Kappa	Strength of agreement	Value of Kappa	Strength of agreement
<0.00	Poor	0.00–0.10	Virtually none	< 0.67	Poor agreement
0.00–0.20	Slight	0.11–0.40	Slight		
0.21–0.40	Fair	0.41–0.60	Fair	0.67-0.80	Fair agreement
0.41–0.60	Moderate	0.61–0.80	Moderate		
0.61–0.80	Substantial	0.81–1.00	Substantial	0.80-1.00	Substantial
0.81–1.00	Almost perfect				

organization, and decision-making in one study²³. Videos can also be a successful tool in evaluating the assessment skills of clinical teachers.²⁴

Table II: Paired ICC between the three Ratings

Rater Pairs	No of ratings	ICC	95% CI	p-value
VA1 – VA2	86	0.957	0.933 – 0.972	0.000
O.S – VA1	86	0.973	0.959 – 0.982	0.000
O.S – VA2	86	0.957	0.934 – 0.971	0.000

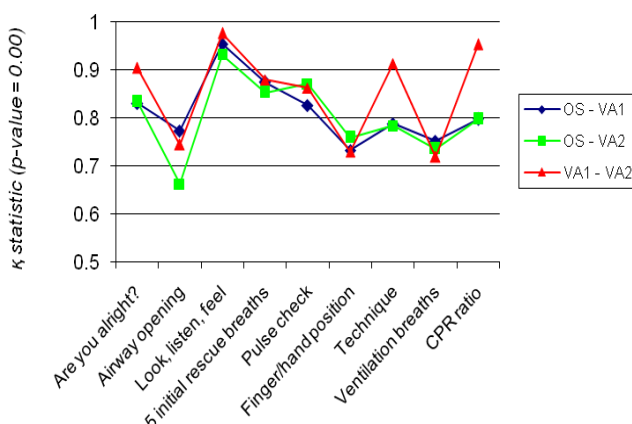
We have used video recordings in the assessment of BLS skills. A similar study addressed the feasibility, reliability, validity, and trainer-trainee agreement in the assessment of tissue-handling skills for surgical trainees using real-time assessment (RTA) and video assessment (VA) and found VA to be highly sensitive.²⁵ In a comparison of retention of CPR skills between a Heart-saver CPR course or a 22-min video training course, CPR skill retention was equivalent in both.²⁶ A study using a six-point descriptive ratings scale for self assessment of basic life support skills in medical students was done at Flinders University in Australia using a Resusci Anne like we did.²⁷

The comparison of two video performances has shown high concordance in terms of ICC as well as categorical Kappa, showing that observation through video is appreciated equally by the observers; Almost two third of the observed tasks have shown high strength of agreement (Figure II; VA1-VA2), however the strength of agreement of the airway opening maneuver, ventilation, and finger or hand position during cardiac compressions is fair to moderate on the krippendorff scale. The correct ventilation depended upon the chest movement of the manikin (the functioning of which had been checked prior to testing) which could not be appreciated clearly through the static video recording in some cases. The same was true for airway opening and cardiac compressions, implying that in order to make videography more effective a tool to assess observer based assessments, recordings through more than one angle and the use of zooming function can be considered to overcome the limitations of focused videography. This is important to consider as video based assessments can provide a valuable tool for distant learning programmes.

The paired analyses of both video based observations with the observed assessment has shown high concordance in terms of ICC, but fair to moderate in terms of assessing the performance as successful or unsuccessful resuscitation. The analyses of individual tasks shows that areas of least concordance are same

for both paired analyses (Figure II). This implies that it might be the limitation of the video based observation that would have led to such discordance; the same task assessments in the VA1-VA2 paired analyses have shown a high strength of agreement. We recommend enhancing the visibility of the interventions recorded by the video camera. The manikin used in these setting needs to have electronic indicators for the assessment of adequate chest movement. Explicit criteria for scoring an item correct need to be described on an instruction sheet for raters. Besides giving clear instructions to new raters, regular rater meetings need to be planned to ensure that criteria are adequately followed. The importance of an objective measurement of resuscitation skills lies in the assumption that, although never proven, adequate life-support skills in a training situation correlates with adequate skills in a real-life situation.

Figure II: Comparison of k-statistics for the three Paired Analyses.



Conclusion

Good quality video recordings if used appropriately have good inter-rater reliability for assessing basic life support skills. With the virtual learning environment being the vogue, video recordings can be used in medical education for assessment.

References

- 1 Advanced Life Support Group [homepage on the internet]. Manchester: Registered charity. [updated 2007 Jun; cited 2007 Aug 12]. Available from: <http://www.alsg.org>
- 2 Hafeez A, Zafar S, Rifaq F, Sultan MA, Southall DP, Muhammad KB. Achieving health millennium development goals through capacity building of health workers-ESSEMCH Strategy. *Pak Paedr J*. 2006; 30(1):3-9.
- 3 Advanced Life Support Group. Advanced Paediatric Life Support. 4th ed. Oxford: Blackwell Publishing; 2005.

- 4 EMCH Reference Manual. Newcastle-under-Lyme: Child Advocacy International; 2004: 136.
- 5 Uitenbroek, Daan G, Binomial. SISA. 1997. <http://home.clara.net/sisa/correl.htm>. (Accessed 3rd Dec. 2006).
- 6 Shrout PE, Fleiss JL. Interclass Correlations: Uses in Assessing Rater Reliability. *Psychol Bull* 1979; 86(2): 420–428.
- 7 Yafee, R.A. Enhancement of Reliability Analysis: Application of Intraclass Correlations with SPSS/Windows v.8. Statistics and Social Science Group, New York University. New York. 1998. [online] <http://www.nyu.edu/its/statistics/Docs/intracis.html> (Accessed December 15th 2006)
- 8 Portney, L.G., Watkins, M.P. Foundations of Clinical Research. Applications and Practice. Appleton & Lange, Norwalk, Connecticut. 1993.
- 9 Boyer KK, Verma R. Multiple raters in survey-based operations management research: a review and Tutorial. *Production and Operations Management* 2000; 9(2):128-140
- 10 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37
- 11 Rigby, A.S. Statistical Methods in Epidemiology; Towards an understanding of kappa coefficient. *Disability and Rehabilitation*. 2000; 22 (8);339-344
- 12 Siegel, S., Castellan Jr., N.J. Nonparametric Statistics for the Behavioral Sciences. International Edition. McGraw-Hill Book Company New York. 1988
- 13 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159.
- 14 Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998;7:301
- 15 Di Eugenio, B. On the usage of Kappa to evaluate agreement on coding tasks. In LREC2000: Proceedings of the Second International Conference on Language Resources and Evaluation. Athens. 2000
- 16 Bahl R, Murphy DJ, Strachan B. BJOG. Qualitative analysis by interviews and video recordings to establish the components of a skilled low-cavity non-rotational vacuum delivery. 2009 Jan;116(2):319-26.
- 17 Fam Med. 2009 Jan;41(1):6-7; author reply 7. Using web-based video to enhance physical examination skills in medical students. Orientale E Jr, Kosowicz L, Alerte A, Pfeiffer C, Harrington K, Palley J, Brown S, Sapieha-Yanchak T.
- 18 Brimble M. Skills assessment using video analysis in a simulated environment: an evaluation. *Paediatr Nurs*. 2008 Sep;20(7):26-31.
- 19 Mazor KM, Haley HL, Sullivan K, Quirk ME. The video-based test of communication skills: description, development, and preliminary findings. *Teach Learn Med*. 2007 Spring;19(2):162-7. Links
- 20 Sorenson DS, Dieter C. From beginning to end: video-based introductory, instructional, and evaluation applications. *Nurse Educ*. 2005 Jan-Feb;30(1):40-3.
- 21 Simpson D, Helm R, Drewniak T, et.al . Objective Structured Video Examinations (OSVEs) for geriatrics education. *Gerontol Geriatr Educ*. 2006;26(4):7-24.
- 22 Hulsman RL, Mollema ED, Oort FJ, Hoos AM, de Haes JC. Using standardized video cases for assessment of medical communication skills: reliability of an objective structured video examination by computer. *Patient Educ Couns*. 2006 Jan;60(1):24-31.
- 23 Kozol R, Giles M, Voytovich A. The value of videotape in mock oral board examinations. *Curr Surg*. 2004 Sep-Oct;61(5):511-4.
- 24 Kevin J, Downie J, Kendall S. Video as a strategy to evaluate and assist clinical teachers with student assessment. *J Nurses Staff Dev*. 2007 Sep-Oct;23(5):223-8.
- 25 Driscoll PJ, Paisley AM, Paterson-Brown S. Video assessment of basic surgical trainees' operative skills. *Am J Surg*. 2008 Aug;196(2):265-72. Epub 2008 Jun 16.
- 26 Einspruch EL, Lynch B, Aufderheide TP, Nichol G, Becker L. Retention of CPR skills learned in a traditional AHA Heartsaver course versus 30-min video self-training: a controlled randomized study. *Resuscitation*. 2007 Sep;74(3):476-86.
- 27 Vnuk A, Owen H, Plummer J. Assessing proficiency in adult basic life support: student and expert assessment and the impact of video recording. *Med Teach*. 2006 Aug; 28(5):429-34.