Evaluation of Medical Students' Performance in Anaesthesia Using a CAE Med-Link Simulator System

D. Cleave-Hogg, P. Morgan, C. Guest
University of Toronto
Centre for Research in Education at the Toronto Hospital
585 University Avenue, Bell Wing 6-600
Toronto, Ontario M5G 2C4, Canada

Abstract

Evaluation of medical students' clinical performance in the operating room is problematic in Anaesthesia, in that patients' safety is of primary importance and students cannot be given unlimited time or opportunities to demonstrate their abilities.

The Anaesthesia Simulator has the potential to offer a controlled environment in which medical students' performance can be assessed. However, there is a need to demonstrate the reliability and validity of the process before it can be integrated into a summative evaluation.

The system used at the Sunnybrook Hospital replicates a realistic operating room setting, involving a full-size electronically controlled mannequin on which interventions with actual equipment and recognised drug administration can be performed.

All fourth year medical students (n=177) were invited to participate in the evaluation project. 145 agreed and they were informed that the results would not be included in the final evaluation and were asked to sign consent forms. Each student worked through a scenario. Students were given 'patient information' and expected to apply their knowledge, demonstrate the necessary technical skills, and use professional judgement. The sessions were videotaped.

Two faculty members reviewed each tape and graded the student in two ways: 1) using the established 25 point evaluation protocol and 2) providing a global rating on a 5 point scale (unsatisfactory - outstanding). These data were analysed for inter-rater reliability. Correlations were calculated with simulator/written/, simulator /clinical and written/clinical grades.

Results indicated satisfactory inter-rater reliability icc 0.77 improving to icc 0.86 when an outlier pair was omitted for lack of attention to task: a problem that must be addressed in further evaluations.

There was no statistically significant correlation between simulator/ written/, simulator/clinical marks and a correlation of r=.23 (significant at r=.21) between written/clinical marks.

The results are encouraging but further studies are needed to improve the process.

Key words
Simulation, Anaesthesia, evaluation, performance assessment

Introduction
The development of virtual simulator technology has captured the interest of educators in the medical specialty of Anaesthesia. Simulators designed for the medical profession, offer numerous teaching and learning opportunities for medical students, residents and practicing physicians. These simulators also have the potential for explicit assessment of performance of a clinical activity in an operating room and, if proved to be suitable, could be incorporated into summative evaluations. Our study was undertaken to explore the feasibility, reliability and validity of using the Anaesthesia CAE Med-Link Simulator System to evaluate the clinical performance of fourth year medical students at the University of Toronto during their Anaesthesia rotation. Results of student and faculty surveys regarding the feasibility of using the simulator for learning and evaluations have been published elsewhere (Morgan and Cleave-Hogg, in press 2000). In this paper we report the results of inter-rater reliability, and case-scenario validity components of the study.

Evaluation of medical competence is viewed as a complex undertaking (Linn, 1993; Van Der Vleuten, 1995; Swanson, 1987) As Van Der Vleuten (1996) warned:

> The art of assessment is more complex than a
> recipe book of agreed testing technology options

Assessment of clinical performance in Anaesthesia education has been problematic in that assessment of performance in the operating room cannot be standardized for particular events, given the variability of patient cases encountered when a medical student is assigned to an anaesthetist in the operating room. In addition Anaesthesia students cannot always be given opportunities to demonstrate their abilities. First, the supervising anaesthetist cannot place a patient at risk if a student performs at a less than adequate level. Adding to this dilemma is the willingness, or confidence, of particular anaesthetists to allow the assigned student to assist in various procedures. Faculty members' tolerance levels of a novice's abilities differ. Second, the student often needs time to think through the next move in a sequence of patient management tasks; time that is not available in most anaesthesia case management processes. This is a particularly sensitive issue in anaesthesia

where interventions are often calculated in seconds and anaesthetists must act quickly, unhindered by novice assistance. Surgeons also are often unable or unwilling to pause in their work to allow time for student reflection. Another concern in the current method of assessment of performance in the real operating room is the multiplicity of subjective assessments given over the period of the rotation. Each student is placed with an anaesthetist who is scheduled for the day, and over a two-week period the student may be supervised by as many as 10 anaesthetists, all of whom have their own criteria for interpreting the marking guide.

Other specialties in medicine have attempted to overcome some of the evaluation problems by the introduction of Objective Standardized Clinical Evaluation (OSCE). The OSCE, developed by Harden and colleagues (Harden 1975) at the Medical School, University of Dundee has become a basic evaluation tool in Canada as well as other countries, using standardized patients as a component of the evaluation. (Rothman,1995) For obvious reasons, the option of standardized patients does not lend itself to anaesthesia procedures nor indeed to many invasive specialty procedures. Currently, in the Department of Anaesthesia, medical students are assessed by a written examination and observations of their activities as they work with attending anaesthetists in an operating room.

The development of a computer-simulated patient offers educators the opportunity to assess students' clinical performance in a controlled, realistic environment. Devit et al. (1998) demonstrated discriminate validity using students, residents and physicians but concluded that case-scenarios should be designed to reflect an appropriate level of education. Gaba, (1988) found that simulation offered a useful tool for the assessment of anaesthesia nurses and residents in crisis management. However, before simulator assessment can be accepted as an integral part of an evaluation in a medical school program it is necessary to gain confidence that the simulator assessment does test the performance required for medical students. Administrators, educators, teaching faculty and students all need to develop this confidence if simulator evaluations are to achieve successful implementation. To this end we undertook a research project to determine whether the simulator-based assessment is feasible, reliable and valid, thus building a foundation for acceptance.

CAE Med-Link Simulator System

The CAE Med-Link Simulator System provides a realistic operating room experience. It consists of a computer-controlled mannequin with simulation of a wide variety of physiological functions such as heart and breath sounds, pulse and end-tidal carbon dioxide. The mannequin takes the place of the patient on the operating table and is connected to regular monitors and surrounded by operating room equipment. Action takes place in the simulator operating room and personnel that make up the medical team are required to dress appropriately in gowns and masks and act out roles such as the surgeon or

circulating nurse. The ongoing activities are intended to duplicate, as closely as possible, conditions likely to be encountered in a real-life situation. Events are programmed in advance and the simulator reacts in a physiologically and pharmacologically appropriate manner. Administered drugs are recognized by the computer and the mannequin-system responds within the parameters set; that is with regard to age, type of disease, fitness level and so on. One of the major advantages of the simulator is that cases can be programmed to challenge learning at the appropriate level of education. Sessions can be videotaped with clear views of actions and audio replication of a student's comments as the case and its problems are discussed. The actual process can also be adjusted through computer input by an assistant in an adjacent room. A third room is designed for video viewing, discussion and feedback.

Working through a case in the simulation centre gives participants a most realistic alternative to a case in the operating room. It also gives students the freedom to fully demonstrate their capabilities, to try out procedures and use their own judgement without endangering a real patient.

Method

Ethics board approval for the project was obtained from the University of Toronto. At the beginning of the academic year, all fourth year students (n=177) were invited to participate during their anaesthesia rotation. Students were informed that they would receive feedback after their simulator session to aid their learning and that the simulator assessment would not be included in the final evaluation. Written consent was obtained from all participants and they were asked to attend an orientation session planned to help them become familiar with the environment.

Case scenarios were designed to reflect the rotation objectives; these included knowledge, judgement and practice skill domains. Prior to the start of the project, the Anaesthesia Undergraduate Curriculum Committee reviewed these cases and consensus was obtained regarding appropriateness and validity of content and expected level of mastery for each component. Learning objectives included rapid sequence induction, treatment of hypovolemia and management of anaphylaxis and tasks such as bag and mask ventilation, laryngoscopy and endotracheal intubation. Each case scenario handout provided to the student gave a brief overview of the case and details of pertinent history, physical examination and laboratory findings.

On the scheduled simulator day, students attending the session first viewed a 10 minute video that addressed the purpose of the study, the capabilities of the simulator mannequin and an overview of the simulator environment, with particular emphasis on the monitors, equipment and pharmaceuticals available. In the simulation operating room, each student was required to work through a 15 minute case scenario, designed to require the application of specific skills and judgement. Students were given written 'patient information' and expected to and

apply their general knowledge, demonstrate the necessary technical skills, use professional judgment to the case in hand.

A faculty member, following a limited-interactive protocol designed for each case-scenario, role-played a circulating nurse and was on hand to provide assistance appropriate to their role without imposing judgement or instruction. All sessions were videotaped for later assessment.

Prior to the commencement of student participation, supervising faculty members attended a workshop to familiarize themselves with the purpose of the study, their role in the simulation centre and the way in which the supervisory protocol was to be used. The protocol detailed introductory statements, instructions to be given to the students and how to manage situations when students were in difficulty. In this way, the dialogue between student and faculty in the simulation environment was controlled and therefore consistent for each student.

Faculty members (designated as faculty evaluators) who assessed the video performances also attended a workshop to ensure that they understood the use of the detailed evaluation checklist, which was prepared for each scenario. The list provided 25 check points that required demonstration of preoperative assessment, an induction sequence, and management of two intraoperative problems. In addition to completing the evaluation checklist, the evaluators were asked to indicate a global rating of the students' performance (5 point rating: 1= unsatisfactory, 5= outstanding) for each student. (Faculty members often declare that they can intuitively judge a student's ability, though this has not been studied in anaesthesia).

After each scenario had been completed, or the time allotted expired, the videotape was switched off. The supervising faculty member met with the students and provided feedback regarding performance after the sessions; this was not part of the evaluation but was intended to enhance the simulator-based learning for the students.

The students' videotaped sessions were then distributed by group to paired faculty evaluators for assessment. Five pairs of evaluators (n=10) assessed the 130 videos; each video was only assessed by one pair of evaluators. Thus each evaluator independently viewed approximately 25-30 videotapes, completed the related performance checklist and provided a global rating for each student.

Results

Out of the class of 177 students, 1 lacked academic prerequisite, 22 declined to participate, 8 did not attend because of weather-related transit cancellation and 6 were attending interviews. Five students undertaking Anaesthesia electives also participated giving a total of 145 students who completed a simulator session. However, during 5 of the sessions we encountered audio problems due to a

microphone malfunction. These 5 videotapes were not used in our calculations, resulting in 140 students whose simulator performance was assessed. Faculty members who participated as evaluators, each spent approximately 10 hours to complete the video evaluations.

The paired evaluators' data were analyzed for inter-rater reliability. Intra-class correlation coefficients were calculated for 'single' and 'mean two rater' student ratings for each of the two pairs, using a two-way random effects model.

Table 1
Estimated reliability for single and paired ratings
Intra-class correlation coefficients

	Pair #1 n=26	Pair #2 n=26	Pair #3 n=34	Pair #4 n=25	Pair #5 n=29
Single rating	0.73	0.58	0.93	0.84	0.77
X paired rating	0.84	0.74	0.96	0.91	0.87

Analysis of intraclass correlation (icc) indicated satisfactory inter-rater reliability with overall icc 0.77 improving to icc 0.86 when the outlier pair was omitted for lack of attention to task. A problem area that will be discussed below. To achieve a reliability of icc 0.90 an estimated 2.68 raters would be required.

Correlation coefficients were calculated between simulator/ written/, simulator/clinical and written/clinical grades. There was no statistically significant correlation between simulator/ written marks: r 0.23, simulator/clinical marks: r 0.13 and low but significant r 0.24 between written/clinical marks.

Inter-item alphas are shown in Table 2.

Table 2
Inter-item alpha for 6 cases

	Inter item alpha
Case 1	0.74
Case 2	0.43
Case 3	0.24
Case 4	0.34
Case 5	0.01
Case 6	0.50

There was a positive correlation between checklist marks and global ratings based on the videotaped performances: r 0.78.

Students' and faculty survey results (Morgan and Cleave-Hogg; in press 2000) indicated that both groups found the simulator session a valuable learning experience, and, given familiarity with the simulator, would be an acceptable evaluation tool.

Conclusion

The inter-rater reliability data indicated an acceptable confidence level providing the evaluators are prepared to give the assessment their full attention. The overall reliability of the raters involved in this study was lower than those reported in a pilot project (Morgan and Cleave-Hogg, 1999) but the results were skewed by the low reliability of one pair. We were concerned with this and discussed the outcomes with the pair in question. One evaluator admitted that he was stretched for time and was trying to do other things while assessing the video performance. This meant that some of the students' performance events may have been missed or not viewed carefully. It was also possible that students' dialogue was not given credit as the evaluator was also interacting with others as he viewed the videotapes. This attitude, obviously, would not be acceptable in a high-stakes, final evaluation situation. Selection, training and vigilance of evaluators thus becomes a critical issue and indicates that the rating process requires continuous testing for inter-rater reliability.

Item analysis of the assessment checklist for case scenarios indicated that faculty agreement of items to be tested and the expected levels of attainment did not predict satisfactory regression analysis outcomes. A careful review of checklist items is required before re-use and new case checklists must be rigorously tested before we can be confident in checklist standadization.

Low correlation between other methods of evaluation was not unexpected because the simulator sessions were testing performance and judgement as well as knowledge. To paraphrase Argyris and Schön,(1974) espoused knowledge and knowledge in action (or applied knowledge) often result in divers outcomes. Ideally, students' "knowledge in action" would be assessed in the regular operating room using a standardized format, but, in reality, patient safety issues and medico-legal considerations do not permit this. We found that anecdotal comments from the students' surveys supported the notion that the simulator offers a viable educational and evaluation tool: the experience of working through a case in the simulation centre was perceived as a valuable experience even though it was more intense and offered greater challenges.

Our current study was limited in that each student worked through only one case. During the actual rotation, we were unable to schedule the necessary time for each student to undertake multiple cases. Comparison across cases is required to ensure greater confidence levels and this element will be integrated into our next study. We are also presently analyzing current evaluation methods in greater depth to ensure that the simulator does indeed offer a more rigorous and

fair assessment of performance and will be a valuable asset to medical evaluations.

Acknowledgments

This study was supported in part with a grant from the Physicians' Services Incorporation Foundation, Canada.
We wish to acknowledge the cooperation of participating students and faculty at the Department of Anaesthesia.

References

Argyris, C. and Schön, D. (1974) *Theory in Practice* San Francisco: Jossey Bass

Devitt, J.H. Kurrek, M. Cohen M.M. *Can medical students be evaluated by a simulator based evaluation tool developed for practicing anesthesiologists? (Abstract)* Anesthesiology 87: A947

Harden, R.M. and Gleeson, F.A. (1975) *Assessment of clinical competence using objective structured examination.* Br. Med. J. 1:447-451

Linn, R.L. (1993) *Education assessment: expanded expectations and challenges*. Educ. Eval. Policy Anal. 15:1-16

Morgan P and Cleave-Hogg, D. (in press 2000) *A Canadian simulation experience: faculty and student opinions of a performance evaluation study*. British Journal of Anaesthesia

Morgan, P. and Cleave-Hogg, D. (2000) Evaluation of medical students' performance using the anaesthesia simulator. Medical Education 34:42-45

Rothman, AI, Cohen,R. (1995) *Understanding the Objective Structured Clinical Examination.: Issues and options.* Annals of Royal College of Physicians and Surgeons of Canada 28:5, 283-287

Swanson, D. Norcini, J. Grosso,L. (1987) *Assessment of clinical competence.* Assess.Eval.Educ. 12:220-246

Tome, JA, Fletcher,J, Lydell DR (1997) *Performance assessment of medical students educated in a simulator environment* (Abstract) Anesthesiology 87: A946

Van Der Vleuten, C.P. Newble, D.I. (1995) *How can clinical reasoning be tested?* Lancet 354:1032-1034

Van Der Vleuten, C.P. (1996) *The assessment of professional competence: developments, research and practical implications*. Adv. Health Science Educ. 1:41-67