

Full scale computer simulators in anesthesia training and evaluation

[Des simulateurs informatisés grandeur nature pour la formation et l'évaluation en anesthésie]

Anne K. Wong MD FRCPC

Purpose: With the advent of competency-based curriculum, technology such as full scale computer simulators have acquired an increasingly important role in anesthesia both in training and evaluation. This article reviews the current role of full scale computer simulators in teaching and evaluation in anesthesia.

Source: This review draws from existing anesthesia and medical education literature in order to examine and assess the current role of full scale computer simulators in anesthesia education today.

Principal findings: The last decade has witnessed a major increase in the use of full scale computer simulators in anesthesia. Many applications have been found for these simulators including teaching and training, evaluation and research.

Despite the increasing use and application of full scale computers in anesthesia in the area of teaching and training, definitive studies evaluating its cost effectiveness, its efficacy compared to traditional training methods or its impact on patient outcome are still pending. Although there is some preliminary evidence of reliability and validity in using the simulator to evaluate clinical competence, development in this area has not progressed enough to justify its use in formal, summative evaluation of competence in anesthesia at this time.

Conclusions: As technology acquires an increasingly important role in medical education, full scale computer simulators represent an exciting potential in anesthesia. However, the full potential and role of simulators in anesthesia is still in development and will require a dovetailing of clinical theory and practice with current research in medical education.

Objectif: Avec l'avènement du programme d'enseignement axé sur les compétences, une technologie comme celle des simulateurs informatisés grandeur nature a acquis un rôle sans cesse plus important pour la formation et l'évaluation en anesthésie. Le présent article passe en revue le rôle actuel de ces simulateurs en anesthésie.

Source : Nous avons extrait des publications existantes sur la formation en anesthésie et en médecine les données qui nous ont permis de vérifier et d'évaluer le rôle actuel des simulateurs.

Constatations principales : Au cours de la dernière décennie, l'usage des simulateurs informatisés grandeur nature s'est grandement développée en anesthésie. Ces simulateurs ont été trouvés utiles dans l'enseignement et la formation, l'évaluation et la recherche.

Malgré leur usage croissant et la multiplicité de leurs applications dans l'enseignement et la formation en anesthésie, des études formelles qui évaluent la rentabilité des simulateurs, leur efficacité comparée aux méthodes de formation traditionnelles ou leur impact sur l'évolution des patients n'ont pas encore été réalisées. Une sorte de fiabilité et de validité préalables à leur usage ont été établies pour évaluer la compétence clinique, mais c'est un domaine qui n'a pas suffisamment évolué pour justifier l'usage actuel de cette technologie dans l'examen formel et le bilan de la compétence en anesthésie.

Conclusion : La technologie joue un rôle croissant dans la formation médicale. Les simulateurs informatisés grandeur nature offrent donc de grandes possibilités en anesthésie. Leur développement constant va nécessiter une harmonisation de la théorie et de la pratique cliniques avec la recherche courante en formation médicale.

THE landscape of medical education is changing dramatically, shifting from what Carraccio *et al.*¹ has termed a process and structure-based curriculum to what is known as an "outcome" or "competency-based curriculum." The former curriculum determines learning on the basis of exposure to specified content over a certain period of time, whereas the latter determines it

From the Department of Anaesthesia, McMaster University, St. Joseph's Healthcare, Hamilton, Ontario, Canada.

Address correspondence to: Dr. Anne K. Wong, Department of Anaesthesia, St. Joseph's Healthcare, 50 Charlton Avenue East, Hamilton, Ontario L8N 4A6, Canada. Phone: 905-522-4941, ext. 3853; Fax: 905-521-6019; E-mail: wongan@mcmaster.ca

This review was self funded and written as part of a course requirement toward a Master of Education degree at the Ontario Institute For Studies in Education of the University of Toronto.

Accepted for publication May 20, 2003.

Revision accepted February 14, 2004.

on the basis of attainment of preset objectives or competencies.¹ Implicit in the competency-based curriculum is that the attainment of these competencies be demonstrable.

In postgraduate specialty training, the shift to a competency-based curriculum was motivated in part by an increasing demand for public accountability of the profession in both Canada and the United States. The Royal College of Physicians and Surgeons of Canada (CanMeds 2000 Project)² has recently defined the key competencies required by a specialist physician as: medical expert, communicator, collaborator, manager, health advocate, scholar and professional. In the United States, the Accreditation Council for Graduate Medical Education has defined these analogously as: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism and systems-based practice.¹

To meet the challenge of this changing landscape, technology has been cited as a major component of the new millennium in medical education.³ In fact, one author predicts that in the future, physicians will acquire competency in invasive procedures with the aid of computer technology before coming into actual contact with a live patient.⁴ The use of computer simulation has increased exponentially, especially in anesthesia, which has leant itself well to the interface of computer technology. In particular, the last decade has seen a major increase in the use of full scale computer simulators in anesthesia.

Full scale computer simulators recreate a realistic work environment which closely mimics real life.⁵ The simulations take place in a fully equipped operating room, complete with an anesthetic machine, monitoring devices, as well as a computer-controlled mannequin on the operating table. The anesthesiologist's or the operating room team's response to the simulated scenarios can be observed, videotaped and later reviewed.

The sophisticated simulation centres have a significant cost both in terms of initial set up as well as ongoing maintenance costs of running the centre. In 1996, Kurrek and Devitt⁶ estimated the construction cost to be \$665,000 CDN and yearly net costs to be \$167,250. Despite the costs involved, a recent study has shown that in just three years, the number of simulation centres has increased from 19 centres worldwide in 1997 to 207 in 2000.⁷

What is the role of full scale computer simulators in anesthesia today? Are they an example of an important advance in anesthesia education or are they a very expensive "technological toy" whose worth is yet to be proven? In particular, what role do they play in

training and evaluation of clinical competency in anesthesia in the new "paradigm shift" of competency-based curriculum?

This article examines these issues by first briefly looking at the historical development of computer simulators to provide the context for their use in anesthesia today. It will then examine the recent literature on the current role that full scale computer simulators play in education and evaluation in anesthesia. The term "simulator" will be used in this article to refer to full scale simulators unless otherwise specified.

History of computer simulation use in anesthesia

Much of the impetus for the development of simulators in anesthesia came from the observed parallels between the environment that the anesthesiologist works in and that found in the aviation industry. The aviation industry has a long history of using simulators for training and maintenance of skills, particularly in crisis management. The first aircraft simulator was built by Edwin Link in 1929. The current aircraft simulators are so realistic that pilots can be trained and certified to fly entirely on the aircraft simulator.⁵ These simulators address the importance of repeated practice, in order for responses to become automatic in the event of an emergency.⁸

Similar to the aviation environment, anesthesia practice in the operating room involves multiple tasks requiring a high degree of vigilance, procedural, monitoring and decision-making skills in a dynamic, complex environment which is affected by the simultaneous interactions of the different members of the operative team. As in aviation, critical life-threatening incidents are rare, but when they do occur, they have potentially disastrous consequences unless the anesthesiologist is able to quickly diagnose and correct the problem. Simulation technology provides a potential way of learning and practicing all the skills involved in anesthesia, including crisis management without harm to a real patient.

The first anesthesia simulator, SIM1, was described in 1969 by Denson and Abrahamson.⁵ Its concepts and construction were ahead of its time. It was originally developed as an aid in learning to intubate as well as to induce anesthesia. It consisted of a mannequin comprising an intubatable airway and upper torso and arms. Despite its cutting-edge technology at the time, cost constraints greatly limited its practical use and further development was abandoned.⁹

Computer simulators did not make a meaningful reappearance until the mid-1980's, starting with the computer screen-only versions such as SLEEPER and BODY simulators as well as the anesthesia simulator

consultant (ASC). By this time, technology had progressed to allow the widespread use of powerful personal computers as well as the ability for complex programs of pharmacology and physiology to model appropriate responses to manipulation of clinical data input. These computer screen simulators (or microsimulators) displayed a realistic representation of the patient, the clinical data and the control panels of the work environment on the screen. They were inexpensive, flexible to use, and could be adapted to present a variety of scenarios which test the anesthesiologists' ability to manage cases. SLEEPER and BODY were used mostly to teach pharmacologic and physiologic principles, whereas, ASC was used mostly to simulate crisis management.^{5,9,10}

In 1986, a team at Stanford headed by Gaba and DeAnda developed a full scale simulator called the Comprehensive Anesthesia Simulation Environment (CASE) to specifically study decision-making processes of anesthesiologists during critical events. The team noted many parallels between decision making during critical events in aviation and anesthesia. In aviation, a focused approach called Crew Resource Management (CRM) had been developed to assist pilots in understanding and dealing with the human factors involved in emergency situations. Taking their cue from the CRM model, Gaba, Fish and Howard subsequently refined CASE to be used in the development of the Anesthesia Crisis Resource Management (ACRM) course.¹¹

The highest level of simulation in the aviation model is the "Line Oriented Flight Training" (LOFT) in which all aspects of a flight are simulated and practiced, including the paperwork, air traffic control and the duties of all primary and support staff. Team Oriented Medical Simulation (TOMS), developed in 1994 by Helmreich, Schaefer and colleagues at the University of Basel, is the medical equivalent of LOFT and emphasizes the combined performance of the whole operating team during critical events.¹²

Currently, full scale simulators have evolved in their use beyond their specific use in teaching crisis resources management to broader aspects of anesthesia training and evaluation. The rest of this article will focus on a discussion of the current use of the simulator in anesthesia education and evaluation.

Full scale simulators in teaching and training

The most common use of the simulator is for teaching and training. In a recent worldwide survey, the majority of the respondents (77–85%) indicated that they were using the simulator for teaching purposes.⁷ There are a number of reasons for their popularity in being used for training in anesthesia:⁹

- 1 They simulate a high degree of reality;
- 2 No risk for a real patient;
- 3 Predictable, programmable, standardizable, reproducible scenarios presented;
- 4 Repeated assessments at present times possible;
- 5 Allows for practice on clinically rare scenarios;
- 6 Videotaping allows for review;
- 7 Simulation can be stopped or restarted for teaching.

One of the seminal uses of the full scale simulator in anesthesia is in training and rehearsing for crisis management. Studies in error evolution and critical incidents in anesthesia have shown that human error is a major contributor (up to 80%).^{13,14} As well, many life-threatening anesthetic emergencies such as malignant hyperthermia have an incidence of less than one in 10,000, therefore the opportunity to encounter them in actual clinical practice may not even occur during the course of an anesthesiologist's career.⁵ Moreover, Gaba, Howard, Fish and colleagues^{8,11} have noted that the traditional training of anesthesiologists does not address teaching crisis management in a systematic way. Drawing from the experiences of the aviation industry, Gaba *et al.*⁹ noted that in order to be effective, these skills must be actively taught rather than acquired through reading about it or by "osmosis." Finally, based on Ericsson *et al.*'s work, it has been shown that the crucial factor in developing expertise is the amount of deliberate practice.^{3,15} The simulator allows the opportunity for repeated practice. As a result, full scale simulators have become a major part of the ACRM course.⁸

The success of the use of the simulator for the ACRM has resulted in a broadening of the simulator use in anesthesia.¹⁶ Many centres are using the simulator for medical student and resident teaching.¹⁷⁻¹⁹

In the context of teaching medical students, the full scale simulator has been used to demonstrate basic airway skills, principles of cardiopulmonary physiology, and to provide an experiential introduction to the practice of anesthesia, as well as aspects of critical care. The medical student may often be relegated to the position of being an observer in many clinical situations due to issues of patient safety.¹⁷ The simulator allows for simulated scenarios to be demonstrated and for the medical student to actively participate in experiential learning without patient harm.

For anesthesia residents, the simulator has been used to familiarize new residents to the operating environment, learn the basic principles of giving a general anesthetic and to learn to deal with common and rare intraoperative problems.^{16,17} It has also been used to run ACRM-type scenarios where higher levels

of training such as communication, team and resource management are taught.¹⁶

Despite the enthusiasm and high ratings given by the participants to such training programs and the intuitive usefulness of training with a full scale simulator, the efficacy of such training compared to traditional methods of teaching is difficult to prove.

The simulated setting will never fully recreate real life and the unpredictability in which real patients may respond to various interventions. As well, in a simulated setting, hypervigilance on the part of the participant because of the anticipation of an "impending crisis" or alternately, a "cavalier attitude" because of the knowledge that this is not a real patient, may hamper how well the simulated setting mirrors the way the participant would respond to real life. Because of these limitations, it is difficult to make conclusions on how well performance in the simulator setting will predict performance in real life. Whether training on a full scale simulator makes a significant difference in patient outcome is even more difficult to ascertain.⁹ Because of its substantial cost, full scale simulators also have to answer to the question of their cost effectiveness, as compared to traditional methods of teaching in anesthesia.

There are only a few studies which look at the efficacy of simulator-based training. Chopra *et al.*²⁰ found that simulator based training of participants in the handling of a malignant hyperthermia crisis resulted in better adherence to established guidelines than a group who had not received similar training. In contrast, Morgan *et al.*²¹ compared the efficacy of simulator-assisted *vs* video-assisted teaching of medical students in their performance in the management of three scenarios. They found no difference in the scores of the students in their group. Nyssen *et al.*²² compared the treatment scores and "time to diagnose" of anesthesia residents and anesthesiologists trained on a full scale simulator *vs* a computer screen-based simulator. They found no difference based on the type of simulator used. However, they did admit that crisis management involves more than just the ability to diagnose and carry out the technical tasks – it also involved behavioural aspects such as communication and team management which was not tested in their study. Thus, despite its apparent advantages as a teaching tool, the verdict is still out as to whether the simulator is cost-effective and whether it improves performance to the extent that it leads to better patient outcome.

If the simulator's role in teaching and training is understood as an adjunct rather than a replacement for traditional modes of clinical teaching in anesthesia, it may be more salient to focus research on how it can

best fulfill this role rather than to compare the efficacy of one mode of teaching over the other. For example, identified gaps in the area of crisis management provided the rationale for the seminal use of the simulator.^{8,9,11} In a similar manner, critical scrutiny to identify weaknesses or gaps in undergraduate or post-graduate anesthesia training will allow the appropriate role of simulators to be better defined.

Use of the full scale simulator for evaluation

In contrast to the commonly accepted use of the simulator for training purposes, there is much more controversy over its use for evaluation. Forty-nine percent of the centres surveyed indicated that the simulator should be used for the maintenance of certification but only 15% are actually using it for this reason.⁷

The New Webster's Dictionary defines *competence* as "sufficient ability" and *competent* as "having the necessary qualities or skills."²³ Clinical competence can thus be defined as the necessary skill or sufficient ability for patient care. "Competencies" on the other hand, such as defined by CanMeds 2000, are domains of competence. On the other hand, *performance* can be described as the *actual demonstration of this skill or competence*.²⁴

Miller²⁵ has described a pyramidal framework for the assessment of the clinical competence of a physician. At the bottom of the pyramid is basic factual knowledge, referred to as "knows." The next level is applied knowledge, referred to as "knows how."

These two levels can be variously assessed by written tests such as multiple choice questions, short answer and essay questions and by oral exams. The third and fourth levels are, respectively, performance *in vitro*, that is, "shows how" in a simulated situation; and performance *in vivo*, that is, "does" in the performance of clinical duties.²⁶ Research over the past four decades has concentrated on developing ways to approximate the pinnacle of the pyramid that is the physician's actual performance *in vivo* or what he/she actually "does."^{25,26}

In the new competency-based paradigm described by Carraccio *et al.*,¹ there is evidence that evaluation of clinical competency will place more emphasis on the "shows how" or performance part of the pyramid. According to Carraccio *et al.*,¹ such assessments will mimic the real tasks of the profession, will be based on direct observation and multiple objective measures, and will be criterion-referenced. This emphasis is already seen in the use of the Objective Structured Clinical Exam (OSCE), which has been widely used in such high stakes examinations as the Medical Council of Canada qualifying exam.²⁷ In the Toronto surgery

residency program, the Objective Structured Assessment of Technical Skills (OSATS), based on similar principles, was developed to assess technical performance during the residency training.²⁸

While the CanMeds 2000 document does not specifically refer to performance assessments, it acknowledges the shortcomings of current evaluation methods. It also admits that there is a “great deal of work to be done in implementing useful systems for the assessment of many components of these new roles and the competencies they embrace. In-training assessment requires further research and development and there is opportunity for complementarity using computerized testing.”²

The formal assessment of anesthesia is still very much at the “knows” and “knows how” stage of the pyramid, despite the fact that like surgery, it is a very technical and procedurally-oriented specialty. Currently, in Canada, the competency of the trainee in anesthesia is assessed by in-training global rating scores and by the Royal College examinations. Due to the nature of the practice of anesthesia, more objective measures of performance such as the OSCE have found limited application in anesthesia prior to the advent of simulators.²⁹ Because full scale simulators can realistically and repeatedly replicate complex scenarios in the operating room environment, they have the potential to be an ideal tool to evaluate clinical competence in anesthesia through performance assessments.²⁹⁻³¹ Is there in fact, adequate evidence in the anesthesia literature to support this role for simulators at this time?

In order to examine this question, it is important to outline the pertinent psychometric criteria which should be taken into consideration when designing a test to evaluate clinical competency. According to a framework proposed by Wass *et al.*,²⁶ the first step is to clearly define whether the test is for formative (educational) or summative (pass/fail) purposes. The second step is to plan the test against the learning objectives of the competency being tested (blueprinting). The third step is to determine the test’s reliability and validity. Reliability of a test refers to the extent to which a test will produce a consistent measurement of the attribute or competency in question when applied under different occasions or by different observers.³² Inter-rater reliability and internal consistency are two ways in which the reliability of a test may be measured. Inter-rater reliability is the degree to which different raters of the test agree with each other. Internal consistency is the degree to which different parts of the test give consistent evaluations of the attribute or competency being tested.

Validity refers to the degree the test is actually measuring the competency in question.³² There are several ways of describing the validity of the test. Construct validity is the extent to which the test fulfills expectations with regard to differentiating the novice from the expert. Content validity is based on the extent that the test measures all pertinent aspects of the competency being studied. Face validity is the extent to which the test reproduces what is experienced in real life. Criterion validity is based on the extent that the test reflects the best existing measure or “gold standard” of treatment. The last is particularly problematic for anesthesia as specific “gold standards” of treatment may be hard to define.

Finally, the last step in designing a test of clinical competency, is standard setting in which the mode of scoring and the minimum standard required to pass is determined. The rigour with which an evaluative test fulfills all these criteria depends on the purpose of the test. Summative evaluation in a high stakes context such as the Royal College certification will require much more rigour than formative evaluation used to provide information and feedback on progress in training.

Byrne and Greaves³³ did a literature review of the studies looking at the use of simulators for performance assessment in anesthesia between the years 1980 to 2000. They found only 13 studies and out of these, only four studies specifically addressed the reliability or validity of the assessments used. These studies included: Devitt *et al.*^{34,35} which looked at inter-rater reliability of scoring scripted scenarios, as well as construct validity of two simulator test scenarios consisting of five intraoperative events each, Gaba *et al.*³⁰ which looked at inter-rater reliability of technical and behavioural scoring of teams during simulated crisis management, and Morgan *et al.*,²⁹ which looked at the reliability of medical student testing with the simulator.

Since that review, several other studies addressing validity and reliability of tests of performance have been published. Forrest *et al.*³¹ have published a study looking at the inter-rater reliability and validity of technical performance assessment of novice anesthesiologists using a simulator. Devitt *et al.*³⁶ published a study to further investigate the construct validity of the test developed in his previous study. Morgan *et al.*³⁷ followed up on their pilot study of 2000 in order to further study the validity and reliability of the evaluation of medical students using the simulator. Schwid *et al.*³⁸ carried out a multi-institutional study of 99 residents to evaluate the validity and reliability of simulator testing in four different scenarios using two

checklist scoring systems. Weller *et al.*³⁹ assessed the reliability of using a global scoring scale to evaluate the performance of anesthesiologists in crisis management in a simulator setting.

The degree to which these studies fulfill the psychometric requirements of an evaluative test as previously discussed, has been summarized in the Table. An important addition to this framework, is the need to clearly define the competency or competencies that are being assessed by the test. As Morgan *et al.*³⁷ 2001 study was an expansion of the 2000 pilot study,²⁹ only the 2001 study was presented in the Table.

On the basis of reviewing the aforementioned studies, several observations can be made. All studies proposed to assess "performance," however, there was no consistent agreement on its definition or what is being measured. Devitt *et al.*³⁴ defined performance as "vigilance, interpretation of data, formulation and implementation of a management plan." Gaba *et al.*³⁰ differentiated between technical performance as being "adequacy of actions taken from a medical and technical perspective," and behavioural performance, as being "decision-making and team interaction processes." Forrest *et al.*³¹ referred specifically to the assessment of the technical aspect of performance, while Morgan *et al.*³⁷ tested 'performance' under the categories of: knowledge of pre-op considerations, preparation, induction of anesthesia (which included demonstrating technical skills), and dealing with intra-operative problems.

Many of the studies did not explicitly state whether the purpose of the test was formative or summative. Only three of eight studies^{30,31,37} explicitly designed the tests to reflect either curriculum objectives or established guidelines for handling certain crisis situations.

While most of the studies demonstrated good inter-rater reliabilities, the reliability of testing with the simulator, as demonstrated by internal consistency measures, was generally low. Content and construct validity were determined in a number of studies whereas face validity was only addressed in one study.³⁶

Most of the studies did not set any minimum criterion or "pass" score for satisfactory performance.

Research on the validity and reliability of performance assessments using the simulator for evaluation has been relatively recent. The existing studies in the literature vary in focus and purpose, thus there is lack of a consensus on what is meant by performance assessment. More problematic, however, is the emphasis on the assessment of "performance" as opposed to assessment of the *underlying competency* that is being "performed." In the performance assessment of crisis management, for example, different competencies,

such as knowledge, technical skill and problem solving, data interpretation skills as well as communication skills may be evaluated. Thus, it is not surprising that in a number of the studies of crisis management, internal consistency was low, suggesting more than one competency domain is being assessed. In designing performance assessments in anesthesia, it would be useful to more clearly define which competency or competencies are actually being evaluated.

In addition, internal consistency as measured by inter-problem or inter-case (as opposed to inter-item correlations within a problem) may also be low due to the phenomenon of "content specificity." If performance assessments in anesthesia are in fact, to some degree, assessing problem solving ability, this phenomenon must be taken into consideration in test design. Based on research into the nature of clinical problem solving over the past 30 years, the phenomenon of content specificity has become axiomatic.⁴⁰ "Content specificity" refers to the finding that problem solving skills are not generic or generalizable but are, in fact, specific to the content and the context of the particular problem.⁴⁰⁻⁴² Thus, typically, correlations across problems are very low, of the order of .1 to .3.^{40,42} As a result, in order to compensate for the low inter-problems or intercase correlations, it is necessary to sample a large enough number of problems in order to obtain reliable measurements of a physician's competency in this area.⁴⁰ In this situation, measurement of inter-problems or intercase reliability is more pertinent than inter-item correlations within a problem. Thus, for improving the reliability of performance assessments of problem solving, one can argue that the more important strategy is to increase the number of problems presented to the examinee rather than eliminate those items that decreased the inter-item reliability coefficient.

The purpose of testing using the simulator also needs to be made explicit and the design and blueprinting of the test must reflect this. Formative testing must be distinguished from summative testing. Studies which look at evaluating team performance need to be clearly distinguished between those that look at individual performance. A consensus needs to be arrived at for appropriate scoring and setting of minimal standard of performance.

Schwid *et al.*³⁸ study on assessing residents using the simulator is an example of where a focused effort to develop the simulator for formative and summative evaluation of clinical competence in anesthesia holds potential. Residents in anesthesia are a specific group whose attainment of clinical competency is expected to improve over time. In the Canadian context, the

TABLE Summary of studies on the use of the full scale simulator for evaluation

Study	Definition of performance assessment (or what is being measured)	Purpose of test summative/formative	Blueprinting	Reliability measures	Validity measures	Standard setting (minimum standard of performance defined)
Morgan <i>et al.</i> 2001	Preoperative consideration Preparation Induction of anesthesia Intraoperative problems	Formative	Yes	Inter-rater reliability Internal consistency	Content validity Convergent validity	No
Forrest <i>et al.</i> 2002	Technical performance	Formative	Yes	Inter-rater reliability	Content validity Construct validity	No
Schwid <i>et al.</i> 2002	Not explicitly stated Response to critical incidents	Formative	Not explicitly stated	Inter-rater reliability Internal consistency	Construct validity Criterion validity	No
Weller <i>et al.</i> 2003	Medical knowledge Behaviour	Not explicitly stated	Not explicitly stated	Inter-rater reliability	Construct validity Content validity Face validity Not assessed	Yes No
Devitt <i>et al.</i> 1997	Problem recognition Formulation of medical diagnosis Institution of Tx	Not explicitly stated	Not explicitly stated	Inter-rater reliability	Construct validity Content validity Face validity Not assessed	No
Devitt <i>et al.</i> 1998	Problem recognition Formulation of medical diagnosis Institution of Tx	Not explicitly stated	Not explicitly stated	Internal consistency	Construct validity	No
Devitt <i>et al.</i> 2001	Problem recognition Formulation of medical diagnosis Institution of Tx	Not explicitly stated	Not explicitly stated	Internal consistency	Construct validity Face validity	No
Gaba <i>et al.</i> 1998	Implementing appropriate technical actions (technical performance) Appropriate crisis management behaviour (behavioural performance)	Not explicitly stated	Not stated for technical performance Blueprinting for behavioural performance	Inter-rater reliability	Not assessed	Yes

Tx = treatment.

Royal College has defined objectives and goals for anesthesia residency programs. Blueprinting can be based on these objectives. Reliability and validity studies as well as standard setting can thus proceed from these starting points. Development in this area can be considered analogous to the development of the OSCE examinations for the Medical Council of Canada qualifying exams and will require a similarly rigorous, focused approach.

However, as with training, the simulator's limitations must be taken into consideration. As Swanson *et al.*⁴³ remarked, "No matter how realistic a performance-based assessment is, it is still a simulation, and examinees do not behave in the same way they would in real life." As in the case with its use in training, the simulator has the potential to be an adjunct to the other modalities currently in use to assess clinical competence. However, based on the review of the current literature, the area of performance assessment in anesthesia still needs further clarification, definition, focus and development. While use of the simulator for assessing progress in training for education purposes (formative evaluation) may have some preliminary evidence of reliability and validity, there is not currently enough development to justify its use for summative evaluation at this time.

Conclusion

The landscape of medical education is undergoing a profound change in the new millennium. The current climate of increasing demand for public accountability and the paradigm shift to competency-based curriculum mean that increased emphasis will be placed on more objective, performance-oriented tests of clinical competence. The role of computer technology, especially in the area of computer simulation is a major part of this changing landscape. It has already spread to the other specialties such as surgery, cardiology³ as well as in emergency and critical care.

However, the dazzle of such technology must not preclude a sober, objective look at its strengths and limitations and what role it should ultimately play in medical education. There are some who point to the example of its use in the airline industry as evidence of its utility.³ On the other hand, there are those who express concern about the dehumanizing effect of an increased reliance on simulators for training³ and still others urge caution in overgeneralizing the parallels between the use of simulators in industry in order to substantiate its usefulness in anesthesia.⁴⁴

In anesthesia, the use of full scale simulators has increased significantly over the last decade. It was originally developed for use in the area of crisis management

in a similar manner that simulators have been used in the aviation industry. In this area, simulator technology has allowed anesthesiologists to fill gaps in traditional anesthesia training, by allowing repeated practice of handling problems that occur infrequently, but need expert intervention for a successful outcome. It also allows for practicing leadership and communication skills in a team-oriented approach. Videotape recordings, and direct feedback add to its utility. Despite the lack of hard data proving that simulator training improves performance over and above the more traditional methods, the full scale simulator has found many applications in teaching and training which span from teaching basic pharmacology and physiology to medical students to demonstrating the general principles of anesthesia to residents as well as crisis management. It provides a unique mode of experiential learning in a realistic, safe, controlled environment that does not expose real patients to harm.

In the area of evaluation, the use of the full scale simulator is much more contentious. Ironically, it may have been the potential usefulness of the simulator as a performance assessment tool that has highlighted the realization of how little work has been done in the area of performance assessment in anesthesia. Work in this area has yet to catch up with the capabilities of the technological tool. Clear definitions and what aspects of performance in anesthesia are to be measured are needed. As "performance" is the active demonstration of the competency, "performance assessments" will need to be better defined and measured according to what skills or competencies are being tested or demonstrated. Although there are studies which show that using the simulator for performance assessments have evidence of reliability and validity for this purpose, it is still premature to suggest that it can be used for high stakes, summative evaluation of clinical competence in anesthesia.

A major difficulty in the development of performance assessments in anesthesia is the difficulty in setting 'gold standard' criteria on the management of certain aspects of anesthesia care as certain outcomes can be arrived at in more than one way.⁴⁵ Adding to the difficulty is the dynamic nature of intraoperative care. As Kapur *et al.*⁴⁶ has pointed out:

"...the time courses of the interactions between patient and clinician are very rapid. Therapeutic approaches to the 'best treatment' may be adjusted midstream if original premises regarding problem etiology are not borne out... For simplicity, most simulator scenarios are designed to highlight only one problem at a time. For training purposes, this approach has merit... however, in an evaluative situa-

tion, this scenario design paradigm could conceivably lead to the test candidate being perfectly able to solve isolated problems in the simulator environment, yet not being competent in the real world of multifaceted, overlapping, simultaneous clinical derangements.”

What is the future of the simulator technology in training and evaluation in anesthesia education? There are some who argue that what we need is not new methods of measuring competency but better methods of recording our observations when we observe our trainees in daily clinical practice.⁴⁷ However, technology is an inseparable part of today’s world as well as the changing landscape of medical education. There is reason to believe that it will continue to play an increasingly important role in the future.

The current climate of competency-based curriculum must not obscure the fact that medical training and evaluation of competency are not an exact science and that a variety of teaching modalities and assessment techniques are required to capture their many aspects. In this regard, the full scale simulator represents an exciting development in anesthesia. However, it must be understood as an adjunct, not a replacement for experiential learning and evaluation in real life. Development of its use in this area requires a dovetailing of clinical theory and practice with current research in medical education and evaluation of clinical competence. The role of full scale simulators in anesthesia is still being mapped out and will depend on a clear understanding of the simulator’s strengths and limitations in anesthesia education.

Acknowledgements

I wish to thank Dr. A. Rothman and Dr. G. Norman for their helpful comments on this article. I also wish to thank Ms. Angela Paparo for her secretarial assistance in the preparation of this manuscript.

References

- 1 Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Acad Med* 2002; 77: 361–7.
- 2 *The Royal College of Physicians and Surgeons of Canada’s Canadian Medical Educational Directions for Specialists 2000 project*. Skills for the new millennium: report of the societal needs working group. 1996; available from URL: http://rcpsc.medical.org/english/publications/canmed_e.html.
- 3 Isenberg SB, McGaghie WC, Hart IR, et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999; 282: 861–6.
- 4 Iserson K. Simulating our future: real changes in medical education. *Acad Med* 1999; 74: 752–4.
- 5 Chopra V. Anaesthesia simulators. *Bailliere’s Clinical Anesthesiology* 1996; 10: 297–315.
- 6 Kurrek MM, Devitt JH. The cost for construction and operation of a simulation centre. *Can J Anaesth* 1997; 44: 1191–5.
- 7 Morgan PJ, Cleave-Hogg D. A worldwide survey of the use of simulation in anesthesia. *Can J Anesth* 2002; 49: 659–62.
- 8 Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviat Space Environ Med* 1992; 63: 763–70.
- 9 Gaba DM. Human work environment and simulators. In: Miller RD (Ed.). *Anesthesia*, 5th ed. Philadelphia: Churchill Livingstone Inc.; 2000: 2613–68.
- 10 Chopra V. What can you do with a simulator? Quality assurance. In: Henson LC, Lee AC (Eds). *Simulators in Anesthesiology Education*. New York: Plenum Press; 1988: 39–48.
- 11 Gaba DM, Fish KJ, Howard SK. *Crisis Management in Anesthesiology*. New York: Churchill Livingstone Inc.; 1994.
- 12 Helmreich RL, Schaefer HG. Turning silk purses into sows’ ears: human factors in medicine. In: Henson LC, Lee AC (Eds). *Simulators in Anesthesiology Education*. New York: Plenum Press; 1998: 1–7.
- 13 Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002; 88: 418–29.
- 14 Cooper JB, Gaba DM. A strategy for preventing anesthesia accidents. *Int Anesthesiol Clin* 1989; 27: 148–52.
- 15 Ericsson KA, Charness N. Expert performance. *Am Psychol* 1994; 49: 725–47.
- 16 Byrick RJ. Intergrating realistic simulation into educational programs: a Canadian perspective. In: Ikeda K, Doi M, Kazama T (Eds). *State-of-the-Art Technology in Anesthesia and Intensive Care*. Amsterdam, Netherlands: Elsevier Science BV; 1998: 175–82.
- 17 Lee AC. Using simulators for medical students and anesthesia resident education. In: Henson LC, Lee AC (Eds). *Simulators in Anesthesiology Education*. New York: Plenum Press; 1998: 23–8.
- 18 Cleave-Hogg D, Morgan PJ. Experiential learning in an anaesthesia simulation centre: analysis of students’ comments. *Med Teach* 2002; 24: 23–6.
- 19 Fish MP, Flanagan B. Incorporation of a realistic anesthesia simulator into an anesthesia clerkship. In: Henson LC, Lee AC (Eds). *Simulators in Anesthesiology Education*. New York: Plenum Press; 1998: 115–9.
- 20 Chopra V, Gesink BJ, de Jong J, Bovill JG, Spierdijk J, Brand R. Does training on an anaesthesia simulator

- lead to improvement in performance? *Br J Anaesth* 1994; 73: 293–7.
- 21 *Morgan PJ, Cleave-Hogg D, McIlroy J, Devitt JH.* A comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology* 2002; 96: 10–6.
 - 22 *Nyssen AS, Larbuissou R, Janssens M, Pendeville P, Mayne A.* A comparison of the training value of two types of anesthesia simulators: computer screen-based and mannequin-based simulators. *Anesth Analg* 2002; 94: 1560–5.
 - 23 *Cayne BS.* *New Webster's Dictionary and Thesaurus of the English Language.* Danbury, CT: Lexicon Publications Inc.; 1992: 200.
 - 24 *Norman GR.* Defining competence: a methodological review. *In: Neufeld VR, Norman GR (Eds).* *Assessing Clinical Competence.* New York: Springer Publishing Company; 1989: 5–35.
 - 25 *Miller GE.* The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65: S63–8.
 - 26 *Wass V, Van der Vleuten C, Shatzer J, Jones R.* Assessment of clinical competence. *Lancet* 2001; 357: 945–9.
 - 27 *Holmboe ES, Hawkins RE.* Methods for evaluating the clinical competence of residents in internal medicine: a review. *Ann Intern Med* 1998; 129: 42–8.
 - 28 *Martin JA, Regehr G, Reznick R, et al.* Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84: 273–8.
 - 29 *Morgan PJ, Cleave-Hogg D.* Evaluation of medical students' performance using the anaesthesia simulator. *Med Educ* 2000; 34: 42–5.
 - 30 *Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R.* Assessment of clinical performance during simulated crisis using both technical and behavioral ratings. *Anesthesiology* 1998; 89: 8–18.
 - 31 *Forrest FC, Taylor MA, Postlethwaite K, Aspinall R.* Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *Br J Anaesth* 2002; 88: 338–44.
 - 32 *Neufeld VR.* An introduction to measurement properties. *In: Neufeld VR, Norman GR (Eds).* *Assessing Clinical Competence.* New York: Springer Publishing Company; 1989: 39–50.
 - 33 *Byrne AJ, Greaves JD.* Assessment instruments used during anaesthetic simulation: review of published studies. *Br J Anaesth* 2001; 86: 445–50.
 - 34 *Devitt JH, Kurrek MM, Cohen MM, et al.* Testing the raters: inter-rater reliability of standardized anaesthesia simulator performance. *Can J Anaesth* 1997; 44: 924–8.
 - 35 *Devitt JH, Kurrek MM, Cohen MM, et al.* Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anesth Analg* 1998; 86: 1160–4.
 - 36 *Devitt JH, Kurrek MM, Cohen MM, Cleave-Hogg D.* The validity of performance assessments using simulation. *Anesthesiology* 2001; 95: 36–42.
 - 37 *Morgan PJ, Cleave-Hogg DM, Guest CB, Herold J.* Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Can J Anesth* 2001; 48: 225–33.
 - 38 *Schwid HA, Rooke GA, Carline J, et al.* Evaluation of anesthesia residents using mannequin-based simulation. A multiinstitutional study. *Anesthesiology* 2002; 97: 1434–44.
 - 39 *Weller JM, Bloch M, Young S, et al.* Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *Br J Anaesth* 2003; 90: 43–7.
 - 40 *Eva KW, Neville AJ, Norman GR.* Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med* 1998; 73: S1–5.
 - 41 *Norman GR.* Evaluation of problem-solving ability. *In: Hart I, Harden R, Walton H (Eds).* *Newer Developments in Assessing Clinical Competency.* Montreal: Heal Publishers; 1986: 47–51.
 - 42 *Norman GR.* Striking the balance. *Acad Med* 1994; 69: 209–10.
 - 43 *Swanson DB, Norman GR, Linn RL.* Performance-based assessment: lessons from the health professions. *Educ Res* 1995; 24: 5–11.
 - 44 *Girard M, Drolet P.* Anesthesiology simulators: networking is the key (Editorial). *Can J Anesth* 2002; 49: 647–9.
 - 45 *Helmreich RL, Davies JM.* Anaesthetic simulation and lessons to be learned from aviation. *Can J Anaesth* 1997; 44: 907–12.
 - 46 *Kapur PA, Steadman RH.* Patient simulator competency testing: ready for take off? (Editorial). *Anesth Analg* 1998; 86: 1157–9.
 - 47 *Whitcomb M.* Competency-based graduate medical education? Of course! But how should competency be assessed? *Acad Med* 2002; 77: 359–60.