Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review

BEME Guide No 4

S Barry Issenberg MD, William C McGaghie PhD, Emil R Petrusa PhD, David Lee Gordon MD & Ross J Scalese MD

This review is published in *Medical Teacher*:

Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL & Scalese RJ (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 27(1): In Press

Author Information

S Barry Issenberg, MD, is an associate professor of medicine, assistant dean for research in medical education, and director of educational research and technology at the Center for Research in Medical Education, University of Miami School of Medicine. His special interest is the use of simulation technology in medical education.

William C McGaghie, PhD, is a professor of medical education and professor of preventive medicine at the Northwestern University Feinberg School of Medicine. He has been a medical education scholar for thirty years.

Emil R Petrusa, PhD, is an associate professor of the practice of medical education in the department of surgery, and associate dean for curriculum assessment at the Duke University Medical Center. His special interests are the use of simulation in medical education and educational evaluation.

David Lee Gordon, MD, is a professor of neurology and medicine, director of neurology teaching, and assistant director of the Center for Research in Medical Education at the University of Miami School of Medicine. His special interest is the education of medical students and residents in the care of neurology patients.

Ross J Scalese, MD, is an assistant professor of medicine and assistant director of educational research and technology at the Center for Research in Medical Education, University of Miami School of Medicine. His special interest is the use of innovative simulations in medical education.

Author Contributions

Study concept and design:	Issenberg, McGaghie, Petrusa, Gordon, Scalese
Acquisition of data:	Issenberg
Analysis and interpretation of data:	Issenberg, McGaghie, Petrusa, Gordon, Scalese
Drafting of the manuscript:	McGaghie
Critical revision of the manuscript for important intellectual content:	Issenberg, Petrusa, Gordon, Scalese
Statistical expertise:	McGaghie, Petrusa
Obtaining funding:	Issenberg, Gordon
Study supervision:	Issenberg

Topic Review Group Members

These individuals participated in the research study at various stages of the project:

Paul Barach, MD, MPH,* Donald D Brown, MD,† Gordon Ewy, MD,† Loryn Feinberg, MD,† Joel M Felner, MD,†* Ira Gessner, MD, †* David Lee Gordon, MD,*† S Barry Issenberg, MD,*† William C McGaghie, PhD,*† Rosanna Millos, MD, † Emil R Petrusa, PhD,*† Stewart Pringle, MD, † Ross J Scalese, MD,† Stephen D Small, MD,* Robert A Waugh, MD,† and Amitai Ziv, MD*

* Pilot phase
† Study phase

BEME Information Scientist: Alex Haig

BEME Consultant: Marilyn Hammick

Guide Series Editor: Pat Lilley

Desktop Publishing: Lynn Bell

ISBN: 1-903934-29-X

This BEME Guide is part of the AMEE Guide Series, published by the Association for Medical Education in Europe. Copies may be obtained from:

Association for Medical Education in Europe (AMEE), Tay Park House, 484 Perth Road, Dundee DD2 1LR, Scotland, UK
Tel: +44 (0)1382 631953 Fax: +44 (0)1382 631987 Email: amee@dundee.ac.uk Website: http:\\www.amee.org

Contents

List of figures and tables

Abstract

Review Date: 1969 to 2003, 34 years

Background and Context: Simulations are now in widespread use in medical education and medical personnel evaluation. Outcomes research on the use and effectiveness of simulation technology in medical education is scattered, inconsistent, and varies widely in methodological rigor and substantive focus.

Objectives: Review and synthesize existing evidence in educational science that addresses the question, "What are the features and uses of high-fidelity medical simulations that lead to most effective learning?"

Search Strategy: The search covered five literature databases (ERIC, MEDLINE, PsycINFO, Web of Science, and Timelit) and employed 91 single search terms and concepts and their Boolean combinations. Hand searching, Internet searches, and attention to the "grey literature" were also used. The aim was to perform the most thorough literature search possible of peer reviewed publications and reports in the unpublished literature that have been judged for academic quality.

Inclusion and Exclusion Criteria: Four screening criteria were used to reduce the initial pool of 670 journal articles to a focused set of 109 studies: (a) elimination of review articles in favor of empirical studies; (b) use of a simulator as an educational assessment or intervention with learner outcomes measured quantitatively; (c) comparative research, either experimental or quasi-experimental; and (d) research that involves simulation as an educational intervention.

Data Extraction: Data were extracted systematically from the 109 eligible journal articles by independent coders. Each coder used a standardized data extraction protocol.

Data Synthesis: Qualitative data synthesis and tabular presentation of research methods and outcomes were used. Heterogeneity of research designs, educational interventions, outcome measures, and timeframe precluded data synthesis using meta-analysis.

Headline Results: Coding accuracy for features of the journal articles is high. The extant quality of the published research is generally weak. The weight of the best available evidence suggests that high-fidelity medical simulations facilitate learning under the right conditions. These include the following:

- *Providing feedback* – Fifty-one (47%) of journal articles reported that educational feedback is the most important feature of simulation-based medical education.
- *Repetitive practice* – Forty-three (39%) of journal articles identified repetitive practice as a key feature involving the use of high fidelity simulations in medical education.
- *Curriculum integration* – Twenty-seven (25%) of journal articles cited integration of simulation-based exercises into the standard medical school or postgraduate educational curriculum as an essential feature of their effective use.
- *Range of difficulty level* – Fifteen (14%) of journal articles address the importance of the range of task difficulty level as an important variable in simulation-based medical education.
- *Multiple learning strategies* – Eleven (10%) of journal articles identified the adaptability of high-fidelity simulations to multiple learning strategies as an important factor in their educational effectiveness.
- *Capture clinical variation* – Eleven (10%) of journal articles cited simulators that capture a wide variety of clinical conditions as more useful than those with a narrow range.
- *Controlled environment* – Ten (9%) of journal articles emphasized the importance of using high-fidelity simulations in a controlled environment where learners can make, detect and correct errors without adverse consequences.
- *Individualized learning* – Ten (9%) of journal articles highlighted the importance of having reproducible, standardized, educational experiences where learners are active participants, not passive bystanders.
- *Defined outcomes* – Seven (6%) of journal articles cited the importance of having clearly stated goals with tangible outcome measures that will more likely lead to learners' mastering skills.
- *Simulator validity* – Four (3%) of journal articles provided evidence for the direct correlation of simulation validity with effective learning.

Conclusions: While research in this field needs improvement in terms of rigor and quality, high-fidelity medical simulations are educationally effective and simulation-based education complements medical education in patient care settings.

Key Words: Medical, Simulation, High-Fidelity, Learning

Review Citation: Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL & Scalese RJ (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 27(1): In Press

Context

Simulation in professional education

Simulations are now in widespread use for professional education and personnel evaluation. Simulations include devices, trained persons, lifelike virtual environments, and contrived social situations that mimic problems, events, or conditions that arise in professional encounters. Simulations range in fidelity or realism from high-end virtual cockpit flight simulators used to train pilots and astronauts to inert sandbags used to train Olympic boxers. Here are several examples drawn from an earlier report (McGaghie, 1999).

- "In April 1997 former U.S. President George H.W. Bush voluntarily parachuted to safety from an airplane 12,500 feet above the Arizona desert. This replicated an experience 50 years earlier when Navy pilot Bush was forced to bail out when his torpedo bomber was shot down during World War II. Commenting on the recent experience septuagenarian Bush declared, 'I'm a new man. I go home exhilarated" (Seligman, 1997). This was not a chance event. Bush trained for the 1997 parachute jump using a virtual reality parachute flight simulator which was originally designed to prepare smoke jumpers to fight forest fires (Aviation Week & Space Technology, 1997).
- Medical students at the University of Michigan learn to provide counsel about smoking cessation from work with simulated patient instructors (SPIs). The SPIs are simulated patients who play the role of genuine patients who are basically healthy yet smoke cigarettes habitually. The SPIs give the medical students detailed feedback about the substance and style of the stop smoking message and evaluate student performance rigorously (Eyler *et al.* 1997).
- Assessment Centers are widely used in business and industry to educate and evaluate managers and executives. However, Spencer and Spencer (1993) report that an Assessment Center has been used to evaluate intelligence officers' capacity to withstand stress under dangerous circumstances, which are simulated with much realism.

  "*...in a well-known assessment center where spies were selected for work behind enemy lines, candidates were locked in a small room with one naked light bulb, then slipped a note that told them they had been captured in the middle of the night photographing documents in the enemy's headquarters. A few minutes later, the door was broken down by men dressed as enemy soldiers, who then forcefully interrogated the subject. These exercises test for self-control and influence skills under stress*" (p. 251).

What, exactly, is a simulation? How is the term defined? As stated elsewhere (McGaghie, 1999): "In broad, simple terms a simulation is a person, device, or set of conditions which attempts to present [education and] evaluation problems authentically. The student or trainee is required to respond to the problems as he or she would under natural circumstances. Frequently the trainee receives performance feedback as if he or she were in the real situation. Simulation procedures for evaluation and teaching have several common characteristics:

- Trainees see cues and consequences very much like those in the real environment.
- Trainees can be placed in complex situations.
- Trainees act as they would in the real environment.
- The fidelity (exactness of duplication) of a simulation is never completely isomorphic with the 'real thing.' The reasons are obvious: cost, [limits of engineering technology], avoidance of danger, ethics, psychometric requirements and time constraints.
- Simulations can take many forms. For example, they can be static, as in an anatomical model. Simulations can be automated, using advanced computer technology. Some are individual, prompting solitary performance while others are interactive, involving groups of people. Simulations can be playful or deadly serious. In personnel evaluation settings they can be used for high-stakes, low stakes, or no stakes decisions" (p. 9).

This definition about simulation exercises squares in nearly all respects with that of Thornton and Mueller-Hanson (2004) in their recent book, *Developing Organizational Simulations: A Guide for Practitioners and Students,* who emphasize the importance of using "...trained assessors to observe behavior, classify behavior into the dimensions being assessed, and make judgments about participants' level of proficiency on each dimension being assessed" (p. 5). Other scholarship demonstrates that reliance on trained assessors to provide educational outcome measurements based on observational ratings is subject to many potential sources of bias (Williams *et al.* 2003). Simulation-based competence measures, *grounded in trainee responses rather than ratings by expert observers*, yield highly reliable and valid educational outcome data (Issenberg *et al.* 2000; Millos *et al.* 2003; Pugh & Youngblood, 2002; Schaefer *et al.* 1998).

Simulation technology has a long legacy of use for education and personnel evaluation in a variety of disciplines and professions. Illustrations include flight simulators for pilots and astronauts, war games and training exercises for the military, management games for business executives, and technical operations for nuclear power plant personnel (McGaghie, 1999; Issenberg *et al.* 2001). There is a growing body of evidence that simulation technology provides a safe and effective mechanism to educate and evaluate professional persons in these fields (Tekian *et al.* 1999).

Simulation in medical education

Medical education has placed increased reliance on simulation technology in the last two decades to boost the growth of learner knowledge, provide controlled and safe practice opportunities, and shape the acquisition of young doctors' clinical skills (Fincher & Lewis, 2002; Gaba, 2000; Issenberg *et al.* 1999a). Intellectual and practical advancement of this work stems from a typology (i.e., framework) that sorts and organizes its many parts.

A typology of *simulators* for medical education has been published by Meller (1997). (This contrasts with the broader term, *simulation*, previously defined.) The Meller typology offers a classification scheme to organize elements of medical simulators. Meller states, "The elements of the analysis include:

P_1 = the patient and/or the disease process
P_2 = the procedure, diagnostic test, or equipment being used
P_3 = the physician or paraprofessional
P_4 = the professor or expert practitioner
p = passive element
a = active element
i = interactive element"

Meller (1997) continues, "Each element of the simulator can be passive, active, or interactive. A passive element usually is provided to enhance the setting or 'realism' of the simulator. Active elements change during the simulation in a programmed way. These elements enhance the simulation and can provoke responses from the student. Interactive elements change in response to actions taken by the student or by any other element of the situation. Any simulated element can be substituted for a real one. In most simulations the (P_3) element is 'real' and represents the student.... The four 'P' types allow the [simulation] developer to assess how realistic the simulation must be to achieve its educational goals" (p. 194).

Applications of many forms of simulation technology to medical education are present and growing. Simulations are becoming an integral part of medical education at all levels (Gaba, 2000; Issenberg *et al.* 1999a). At least five factors contribute to the rise of simulations in medical education: (a) problems with clinical teaching; (b) new technologies for diagnosis and management; (c) assessing professional competence; (d) medical errors, patient safety, and team training; and (e) the role of deliberate practice.

Problems with clinical teaching

Changes in the delivery of health care trigger major shifts in medical education methods. For instance, in the United States, the pressures of managed care are shaping the form and frequency of hospitalizations, resulting in higher percentages of acutely ill patients and shorter in-patient stays. This results in less opportunity for medical learners to assess patients with a wide variety of diseases and physical findings. Despite increased cost-efficiency in outpatient care, reductions in physician reimbursement and shrinking financial resources constrain the educational time that physicians in training receive in this environment. Consequently, physicians at all educational levels find it increasingly difficult to keep abreast of skills and topics that frequently appear in practice.

These problems have a direct effect on clinical skills training, such as bedside cardiology. For example, despite evidence that accurate clinical examination of patients with cardiac signs and symptoms is a cost-effective diagnostic modality (Roldan *et al.* 1996), direct bedside teaching of these skills is occurring with decreasing frequency. The result is a decline in the quality of healthcare providers' bedside skills and a reduction in the ability to provide high-quality and cost-effective medical care. The loss of clinical acumen was documented in a recent study that demonstrated house officers have difficulty identifying common cardiac findings. That study also stressed the need for structured, supplemental strategies to improve clinical education, including the use of simulation systems for training (Mangione & Nieman, 1997).

New technologies for diagnosis and management

The advent of new technologies in medicine has revolutionized patient diagnosis and care. The past 30 years have witnessed the development of flexible sigmoidoscopy and bronchoscopy, minimally invasive surgery including laparoscopy, and robotics for orthopedics and cardiology. The benefits of these methods include (a) reduced postoperative pain and suffering, (b) shorter hospitalization and earlier resumption of normal activity, and (c) significant cost savings.

However, the psychomotor and perceptual skills required for these newer techniques differ from traditional approaches. Research indicates that these innovative methods may be associated initially with a higher complication rate than traditional practice (Deziel *et al.* 1993). These newer technologies have created an obstacle to traditional teaching that includes hands-on experience. For example, endoscopy requires guiding one's maneuvers in a three-dimensional environment by watching a twodimensional screen, requiring the operator to compensate for the loss of the binocular depth cue with other depth cues. Simulation technology has been introduced as a method to train and assess individuals in these new techniques. A recent survey of training program directors stressed the importance of virtual reality and computer-based simulations as technological tools in clinical education (Haluck *et al.* 2001).

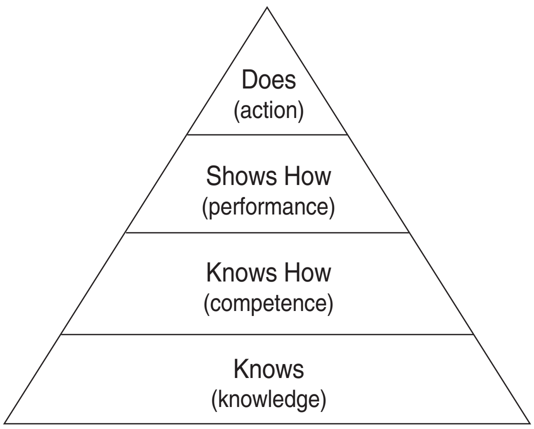
Assessing professional competence

The Accreditation Council for Graduate Medical Education (ACGME) asserts there are six domains of clinical medical competence (ACGME Outcomes Project, 2003). The list of six was published in response to the belief that professional competence should be defined and evaluated in a way that includes all important domains of medical practice. The six domains are:

1 Patient Care
2 Medical Knowledge
3 Practice-based Learning and Improvement
4 Interpersonal and Communication Skills
5 Professionalism
6 Systems-based Practice

For each domain of competence, Miller (1990) earlier proposed a framework that argues there are four levels at which a medical learner should be assessed. The levels (displayed in Figure 1) are: (a) **knows** (knowledge) - recall of facts, principles, and theories; (b) **knows how** (competence) – ability to solve problems and describe procedures; (c) **shows how** (performance) – demonstration of skills in a controlled setting; and (d) **does** (action) – behavior in real practice.

Figure 1: Miller's Framework for clinical assessment. Reprinted with permission from *Academic Medicine*



Simulation technology is increasingly being used to assess the first three levels of learning because of its ability to (a) program and select learner-specific findings, conditions, and scenarios; (b) provide standardized experiences for all examinees; and (c) include outcome measures that yield reliable data (Issenberg *et al.* 2002).

Medical errors, patient safety, and team training

Recent studies and reports, including the U.S. Institute of Medicine's *To Err is Human* (Kohn *et al.* 1999) and a subsequent empirical study reported in the *Journal of the American Medical Association* (Zahn & Miller, 2003), have drawn attention to the perils of healthcare systems worldwide (Barach & Moss, 2002; Brennan *et al.* 1991). These reports have highlighted the tensions between accountability and improvement, the needs of individual patients and benefit to society, and financial goals and patient safety.

Most medical errors result from problems in the *systems* of care rather than from individual mistakes (Bogner, 1994). Traditional medical training has focused on individual learning to care for individual patients. Medical education has neglected the importance of teamwork and the need to develop safe systems (Helmreich & Schaefer, 1994). The knowledge, skills, and attitudes needed for safe practice are not normally acquired, nor are they required, as part of medical education. For more than two decades, non-medical disciplines such as commercial aviation, aeronautics, and the military have emphasized team (crew) resource training to minimize adverse events (Brannick *et al.* 1997). In addition, the Institute of Medicine report asserts, " . . . health care organizations should establish team training programs for personnel in critical care areas . . . using proven methods such as crew resource management techniques employed in aviation, including simulation" (Kohn *et al.* 1999).

Deliberate practice

Instructional science research demonstrates that the acquisition of expertise in clinical medicine and a variety of other fields (e.g., professional sports, aviation, chess, musical performance, academic productivity) is governed by a simple set of principles (Ericsson & Charness, 1994; Ericsson *et al.* 1993; Ericsson & Lehman, 1996). These principles concern the learner's engagement in deliberate practice of desired educational outcomes. Deliberate practice involves (a) repetitive performance of intended cognitive or psychomotor skills in a focused domain, coupled with (b) rigorous skills assessment, that provides learners (c) specific, informative feedback, that results in increasingly (d) better skills performance, in a controlled setting. Scholarly research about the acquisition of expertise consistently shows the importance of intense, deliberate practice in a focused domain, in contrast with so-called innate abilities (e.g., measured intelligence) for the acquisition, demonstration, and maintenance of skills mastery (Ericsson, 2004).

A recent cohort study conducted at five academic medical centers (Duke, Emory, Miami, Mt. Sinai, Northwestern) illustrates the utility of deliberate practice in medical education (Issenberg *et al.* 1999b). Fourth year medical students enrolled in a 4-week cardiology elective received either (a) a 2-week multimedia educational intervention followed by 2 weeks of ward work, or (b) 4 weeks of customary ward work (i.e., teaching rounds, patient workups). The multimedia intervention engaged the medical students in deliberate practice of cardiology bedside skills using 10 computer-based teaching modules linked to the "Harvey" cardiology patient simulator (Issenberg *et al.* 1999a). Both student groups took an objective, multimedia, computer-based pretest and posttest specifically developed to provide reliable measures of cardiology bedside skills (Issenberg *et al.* 2000). Research outcomes show that (a) intervention group performance increased from 47% to 80% after 2 weeks of deliberate practice, and (b) comparison group performance increased from 41% to 46% after 4 weeks of evaluating patients in the hospital and clinic and seeing more patients than students in the intervention group. Medical students in the intervention group that engaged in deliberate practice acquired nearly twice the core bedside cardiology skills, in half the time as the comparison group, with little or no faculty involvement. This research has been replicated in a sample of internal medicine residents with nearly identical results (Issenberg *et al.* 2002).

Another deliberate practice intervention study, a randomized trial with wait-list controls, evaluated acquisition of advanced cardiac life support (ACLS) skills among internal medicine residents using a full-body mannequin simulator. Residents who received the educational intervention performed 38% better on a reliable clinical skills evaluation than residents in the wait-list control group. Following crossover and a second deliberate practice intervention, residents formerly in the wait-list control group surpassed the clinical performance outcomes of the first intervention group (Wayne *et al.*, 2005, in press). Deliberate practice, not just time and experience in clinical settings, is the key to development of medical clinical competence.

Quality in medical education research

Coincident with the expansion of simulation technology in medical education is a growing call for higher quality in medical education research. This call comes from several sources. One source is editors of influential medical journals. For example, Stephen J Lurie, former Senior Editor of the *Journal of the American Medical Association*, recently published an essay titled, "Raising the passing grade for studies of medical education" (Lurie, 2003). Lurie documents many flaws in medical education research and calls for common metrics, increased standardization of educational interventions, better operational definitions of variables, and, at bottom, more quantitative rigor. Lurie's call is echoed by Jerry A. Colliver (2003) Editor of *Teaching and Learning in Medicine: An International Journal.*

A second source calling for higher quality medical education research is a Joint Task Force (2001) of the journal *Academic Medicine* and the GEA-RIME Committee of the Association of American Medical Colleges. A report of this Task Force titled, "Review Criteria for Research Manuscripts" provides detailed technical suggestions about how to improve medical education research and its sequelae, scholarly publications.

A third call for improved medical education research rests within the research community. To illustrate, a team of investigators under auspices of the Campbell Collaboration recently attempted to perform a systematic review of the research evidence on the effectiveness of problem based learning (PBL) in medical education (Newman & the Pilot Review Group, 2003). However, due to the abundance of low-quality studies, heterogeneity of the published investigations, and disagreement about basic research terms and conditions, the systematic research review could not be performed as planned. Thus despite the widespread use of PBL in medical education worldwide there are little systematic, reliable empirical data to endorse its effectiveness as a learning modality. (Of course, the same could be said about the effectiveness of lectures in the basic sciences and clinical disciplines as a source of knowledge acquisition, especially compared with reading.)

Best evidence medical education (BEME)

The Best Evidence Medical Education (BEME) Collaboration (Harden *et al.* 1999) involves an international group of individuals, universities, and organizations (e.g., AMEE, AAMC), committed to moving the medical profession from opinion-based education to evidence-based education. The goal is to provide medical teachers and administrators with the latest findings from scientifically grounded educational research. This permits the teachers and administrators to make informed decisions about the kinds of evidence-based education initiatives that boost learner performance on cognitive, conative, and clinical measures. BEME rejects the medical education legacy that has relied little on evidence in its decision-making, relying instead on pseudoscience, anecdotes, and flawed comparison groups. The BEME philosophy asserts that in no other scientific field are personal experiences relied on to make policy choices and in no other field is the research base so inadequate.

BEME scholarship " ...involves a professional judgment by the teacher [or administrator] about his/her teaching taking into account a number of factors - the **QUESTS** dimensions: the **Quality** of the research evidence available - how reliable is the evidence? the **Utility** of the evidence - can the methods be transferred and adopted without modification, the **Extent** of the evidence, the **Strength** of the evidence, the **Target** or outcomes measured - how valid is the evidence? and the **Setting** or context - how relevant is the evidence?" (Harden *et al.* 1999, p. 553).

The international BEME Collaboration has three broad purposes. First, to produce systematic reviews of medical education research studies that capture the best evidence available and also meet users' needs. Second, to disseminate information worldwide to all stakeholders to make decisions about medical education on grounds of the best available evidence. Third, to create a culture of best evidence medical education among teachers, administrators, educational institutions, and national and international organizations.

This report

This BEME report is one of several outcomes from a project originating from a February 2001 invitation by the BEME Executive Committee to the Center for Research in Medical Education (CRME) at the University of Miami School of Medicine (USA). The University of Miami CRME accepted the charge to review and synthesize existing evidence in educational science that addresses a specific question, "What are the features and uses of high-fidelity medical simulations that lead to effective learning?" This report presents the methodological scope and detail of the project, its principal findings, and a discussion about what the findings mean for evidence-based medical education today and tomorrow.

Three sections follow. First, a **Methods** section describes two research phases: (a) a *pilot phase* that reports preparation steps taken before the research review got underway, and (b) the *study phase* that gives specific details about the bibliometric search strategy and the research review and data synthesis. Second, a **Results** section presents our findings in detail, including descriptive outcomes about research reports included in the systematic review and a list of ten key features of high-fidelity medical education simulations that evidence shows lead to effective learning. Third, a **Conclusions** section that (a) interprets our principal findings, i.e., "What do the findings mean?" (b) acknowledges the limits [not failure] of this and other BEME reviews; (c) critiques the quality and status of current research in the field of high-fidelity simulations in medical education; and (d) calls for a bolder, more rigorous research agenda in this and other domains of medical education internationally.

Methods

Eight step pilot phase

An eight-step pilot phase was undertaken to prepare for the formal, systematic research review. The pilot phase was deliberately cautious, intended to identify and fix research problems before the larger study got underway.

Step 1: *BEME Invitation.* The BEME Executive Committee (R.M. Harden, Chair) invited the Center for Research in Medical Education of the University of Miami School of Medicine in February 2001 to conduct a BEME systematic review addressing a specific question, "What are the features of high-fidelity medical simulations that lead to most effective learning?" The invitation was offered to the Miami Center for two reasons: (a) its expertise [grounded in history and personnel] in the use of simulation technology in medical education, and (b) a track record of performing multi-institutional medical education research studies consonant with the BEME model. The Miami Center agreed to undertake the project under the leadership of S.B. Issenberg, M.D., its Director of Educational Research and Technology.

Step 2: *Formation of the pilot Topic Review Group (TRG).* The second step was to assemble an interdisciplinary group of expert scientists and clinicians to plan and manage the pilot phase of the systematic review. Three criteria were used to select individuals for TRG participation: (a) *international representation*, i.e., experts from a variety of countries worldwide; (b) persons with expertise involving a *wide variety of medical simulations*, e.g., the "Harvey" cardiology patient simulator and simulators used in anesthesiology, surgery, and virtual reality applications; and (c) experts with appropriate knowledge of research methods, educational measurement, and the process of conducting systematic literature reviews.

The pilot phase TRG included representatives from eight medical institutions:

1 Duke University Medical Center (USA)
2 Emory University Medical School (USA)
3 Northwestern University Feinberg School of Medicine (USA)
4 University of Chicago Pritzker School of Medicine (USA)

5 University of Dundee Faculty of Medicine (UK)
6 University of Florida College of Medicine (USA)
7 University of Miami School of Medicine (USA)
8 Tel Aviv University (Israel)

Step 3: *Address conceptual issues.* Two conceptual questions framed and focused the pilot work of the TRG. (a) What is the definition of *effective learning*? and (b) What are the *elements of a high-quality, systematic literature review*?

We dissected *effective learning* into two parts. *Effectiveness* was classified according to an expansion of the four Kirkpatrick (1998) training criteria. [The Kirkpatrick criteria are nearly identical to Miller's (1990) four-level framework for medical learner assessment that was cited earlier.] Effectiveness of medical learning is conceived as an ordinal construct ranging from:

- Level 1—participation in educational experiences
- Level 2_a—change of attitudes
- Level 2_b—change of knowledge and/or skills
- Level 3—behavioral change
- Level 4_a—change in professional practice
- Level 4_b—benefits to patients

The definition of medical *learning* focused on measured educational outcomes having clinical medical utility. We chose nine nominal yet overlapping categories:

- clinical skills
- practical procedures
- patient investigation
- patient management
- health promotion
- communication
- information skills
- integrating basic sciences
- attitudes and decision-making

Our definition of the *elements of a high-quality, systematic literature review* is based on previous work published by Frederic Wolf (2000) in *Medical Teacher.* The eight *elements* shown in Table 1, range from stating the objectives of the review to conducting an exhaustive literature review, tabulating characteristics of eligible studies, synthesizing results of eligible studies, and writing a structured report. Quantitative research synthesis (meta-analysis) is used if appropriate and possible. Not all systematic literature reviews lend themselves to quantitative synthesis (Newman & the Pilot Review Group, 2003).

Table 1: Elements of a high quality systematic review

1	State objectives of the review, and outline eligibility (inclusion/exclusion) criteria for studies
2	Exhaustively search for studies that seem to meet eligibility criteria
3	Tabulate characteristics of each study identified and assess its methodological quality
4	Apply eligibility criteria and justify any exclusions
5	Assemble the most complete dataset feasible, with involvement of investigators
6	Analyze results of eligible studies. Use statistical synthesis of data (meta-analysis) if appropriate and possible
7	Perform sensitivity analyses, if appropriate and possible (including subgroup analyses)
8	Prepare a structured report of the review, stating aims, describing materials and methods, and reporting results

Source: Wolf (2000). Adapted from Chalmers I (1993). The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care (pp. 156-65). In Warren KS, Mosteller F (eds). Doing more good than harm: the evaluation of health care interventions. Annals of the New York Academy of Sciences (Vol 703). Reprinted with permission from *Medical Teacher.*

Step 4: *Defining the research question and search criteria.* The fourth step in the pilot process was refinement of the research question and search criteria. Our TRG received the question, "What are the features of high-fidelity medical simulations that lead to most effective learning?" from the BEME Executive Committee. The question was used to generate literature search criteria. The TRG developed search criteria to define each of the following components of the research question: (a) *features*, (b) *high-fidelity simulators*, and (c) *effective learning*. Examples of the pilot search criteria include:

- *Features*: The fidelity of a simulator by expert opinion. What is simulator validity, i.e., can the simulator in evaluation mode differentiate a novice from an expert? Is there a built-in teaching and assessment system (e.g., Issenberg *et al.* 2000; Millos *et al.* 2003; Pugh & Youngblood, 2003). How are local logistics managed?
- *High-Fidelity Simulator*: There is a distinction between a simulator that changes and responds to the user and a simulator that remains static, e.g., task trainer (Meller, 1997). We assigned three broad categories: (a) realistic, three-dimensional procedural simulators; (b) interactive simulators, e.g., responds to prompts, probes, and procedures; and (c) virtual reality simulators.
- *Effective Learning*: Examples include documented improvement in any of the nine previously defined clinical categories (e.g., clinical skills, health promotion, integrating basic sciences) that capture key medical education outcomes. These learning outcomes were classified according to the modified

[ordinal] Kirkpatrick (1998) criteria (e.g., participation, attitude change, behavior change, benefits to patients).

Step 5: *Literature search.* The next step in the pilot process was the literature search. The TRG agreed that the pilot study should include 30 to 40 data-based research reports without tight constraints about the type of article (e.g., randomized trial, cohort study) or population of learners (e.g., medical students, residents) to obtain a broad, representative sample of published articles. The pilot literature search generated approximately 200 references. An initial screen based on the presence of original data, versus essays and statements of opinion, resulted in 32 studies for the TRG to review.

Step 6: *Early meeting and tryout.* A key step in the pilot process was a Simulation TRG meeting that occurred on June 6-7, 2001 in Miami, Florida (USA). The TRG:

- *Reflected* on the search question and revised it to state, "What are the features **and uses** of high-fidelity medical simulations that lead to effective learning?" The group asserted that the intended use of a simulation is equally important as its specific features.
- *Formulated* pairs of research study coders to work as teams during the pilot phase.
- *Studied* abstracts from the 32 articles to determine which ones should be coded for systematic review. Sixteen studies (50%) were not included because (a) 14 reports did not meet basic criteria (e.g., no high-fidelity simulator, no discussion about simulator use), and (b) two reports were not published in English and could not be translated promptly.
- *Implemented* a coding form provided by the BEME Executive Committee. Four two-person teams coded one article and compared their findings. Later, the full TRG convened to review its findings, clarify unfamiliar terms, and make suggestions about revising the coding sheet.
- *Continued to code* the remaining 15 articles. Each article was reviewed by a team of two TRG members. Each team reviewed the results of their individual coding and gave suggestions for coding improvement.
- *Synthesized* all of the comments and suggestions and authored a revised coding form that was more relevant to simulation-based medical education. The form added items directly pertinent to high-fidelity medical simulations.

After the TRG meeting its convenor (SBI) finalized the coding form and instructions for its use. These were distributed to all TRG members. Also, the TRG leader summarized the findings of the pilot phase which were presented at the BEME Workshop during the Summer 2001 *Association for Medical Education in Europe* (AMEE) meeting in Berlin.

Step 7: *Problems and resolutions.* Five problems arose as a result of the Simulation TRG being one of the first to conduct a pilot review.

- The [local] University of Miami library was late in acquiring two journal articles before the June 2001 TRG meeting. This resulted in other TRG members and the BEME Administration using their own University libraries to obtain articles.
- The coding sheet and description of terms was not provided to the TRG before its meeting. This caused confusion and misunderstanding during the first coding session. Once the TRG practiced with the coding sheet and agreed on terminology, later rounds of coding occurred with less confusion and improved interrater agreement for each article. This was reflected in their comments and also in their coding sheet answers.
- All of the TRG members found the coding sheet inappropriate for narrative review articles. The coding categories did not apply and TRG members considered items on the coding sheet only to later realize they did not apply to a review article. Questions were rearranged to better orient the coder to the type of article (e.g., one of the first questions became research design) to better focus the reviewer for subsequent items on the coding sheet.
- There was no operational Internet database with common access by TRG members. This inhibited the ability of the TRG leader to add citations to the database. Internet access would enable members of the TRG to quickly determine if an abstract was already included in the review process, whether a full article had been obtained, and whether it had been coded.
- Before and after the TRG meeting many of the members were slow to respond to e-mails asking them for comments on a variety of issues. As a result, this meant more work for the TRG leader and less shared input from others.

Step 8: *What worked.* During the pilot study period, there was excellent communication between the TRG leader and an information scientist at the University of Dundee (UK). This facilitated the creation of search criteria and generation of references to be included in the pilot study. The most important aspect of the pilot project was having all members together for two dedicated days to review the topic question and search criteria, orient the members to the coding sheet, and practice coding articles. Finally, the presence of Drs. Ian Hart and Ronald Harden at the June 6-7, 2001 Simulation TRG meeting in Miami, Florida (USA) to answer questions and to provide focus about the broad goals of the BEME project provided objective guidance.

Summary of pilot methodological issues

- The entire pilot process was funded by the Center for Research in Medical Education at the University of Miami School of Medicine. (Over the course of the project, the cost can be significant, especially if TRG meetings occur.)
- To insure a reviewer group with a broad background, we selected individuals with expertise in diverse areas including simulation, medical education, and research methods.
- An important step before the process began was agreement on the question and search criteria. Our TRG elected to adopt the suggested question because we believed it represented what most medical educators would want to know about simulation. The question was modified slightly to include simulation **use** in addition to simulation features.
- Several TRG members were concerned that the coding process would lead to quantitative data that may not answer the review question. These concerns lessened when the QUESTS criteria were suggested as a mechanism to judge articles. In addition, the TRG added items to the end of the coding sheet that sought information to better answer our question.
- It is important to create an accessible Internet database that reflects the current state of the topic review.
- There was concern among TRG members that the coding sheet had not been studied to assess its objectivity in reducing reviewers' background bias. Reviewer training and practice is needed to reduce rater bias and to boost inter-rater agreement and objective coding.

Conclusions relating to pilot process

All of our TRG members believe the pilot process was a valuable learning experience and suggest that other TRGs undergo a similar exercise before engaging in a full BEME review. Topic group leaders should be fully informed and experienced with the coding sheet and instructed to educate other group members. It is important to provide a meeting of TRG members to orient themselves to the search questions, coding process, and other study features. While a dedicated meeting may not be feasible, there may be other opportunities to convene at national or international medical education meetings (e.g., AMEE, AAMC, ASME, and Ottawa Conference).

Our pilot study did not include enough articles to enable us to answer our original question. It did allow our TRG to become familiar with the process and to appreciate the considerable time and effort needed to insure its completion. We suggest that empirical reports should be chosen that have measurable outcomes explicitly stated and studied. Review and descriptive articles are tedious and difficult to assess when grouped together with randomized trials, cohort studies, and case-control studies. Our TRG has elected to separate review articles and provide an annotated qualitative list of its own.

The results of the pilot phase are in close agreement with the "Twelve tips for undertaking a systematic review" discussed in an article published in *Medical Teacher* (Reeves *et al.* 2002). Future BEME TRGs will benefit by attending to our experience and to the advice from Reeves and his colleagues.

Six step study phase

The final implementation phase involving the **Methods** of the systematic review was performed by the BEME medical simulations TRG in six steps. The six steps were: (a) identify the final cohort of BEME research coders; (b) BEME research coder training; (c) literature search strategy; (d) research study selection; (e) data extraction and coding; and, (f) data analysis and synthesis.

Step 1: *Final cohort of BEME research coders.* The final cohort of research study coders included the authors of this report (Issenberg, McGaghie, Gordon, Petrusa, Scalese) and eight other Working Group Members (Brown, Ewy, Feinberg, Felner, Gessner, Millos, Pringle, Waugh). All of these individuals participated in the project without compensation or other incentives.

Step 2: *BEME research coder training.* The BEME research coders received one session of *frame of reference training* adapting performance appraisal procedures described by Woehr and Huffcutt (1994). This involved orienting the coders to key features of the published research studies (i.e., research design, measurement methods, data analysis), seeking consensus about the key features from discussion and feedback, and judging the key features using a uniform set of quality standards embedded in the coding sheet. The research coder group analyzed a single, illustrative study together to reach agreement about terminology, key features, and quality standards. Independent research study coding began immediately after the training session.

Step 3: *Literature search strategy.* Medical education, and professional literatures about the features and use of high-fidelity medical simulations that lead to most effective learning were searched systematically in collaboration with experienced reference librarians. The purpose of the search was to identify relevant studies that document the impact of high-fidelity medical simulations on key learning outcomes. Databases were targeted that would yield reports of original research in this area.

The search timeframe spanned 34 years from June 1969 when the seminal article about simulation in medical education was published by Abrahamson *et al*. (1969) to June 2003. The search covered five literature databases (ERIC, MEDLINE, PsycINFO, Web of Science, and Timelit) and employed a total of 91 single search terms and concepts, and their Boolean combinations (Table 2). We also hand searched key publications that focused on medical education or were known to contain articles on the use of simulation in medical education. These journals included *Academic Medicine, Medical Education, Medical Teacher, Teaching and Learning in Medicine, Surgical Endoscopy,* and *Anesthesia and Analgesia*. In addition, we also hand searched the annual *Proceedings* of the *Medicine Meets Virtual Reality Conference* and the biannual *Ottawa Conference on Medical Education and Assessment.* These *Proceedings* include “grey literature” (e.g., papers presented at professional meetings, doctoral dissertations) determined by our TRG to contain the most relevant references related to our review. Several basic Internet searches were also done using the Google.com search engine. The aim was to perform the most thorough literature search possible of peer reviewed publications and reports in the unpublished “grey literature” that have been judged for academic quality.

All of the 91 search terms could not be used within each of the five databases because the databases do not have a consistent vocabulary. Each database also has unique coverage and emphasis. Attempts were made to use similar text word or keyword/phrase

Figure 2: Literature review and selection of articles for review

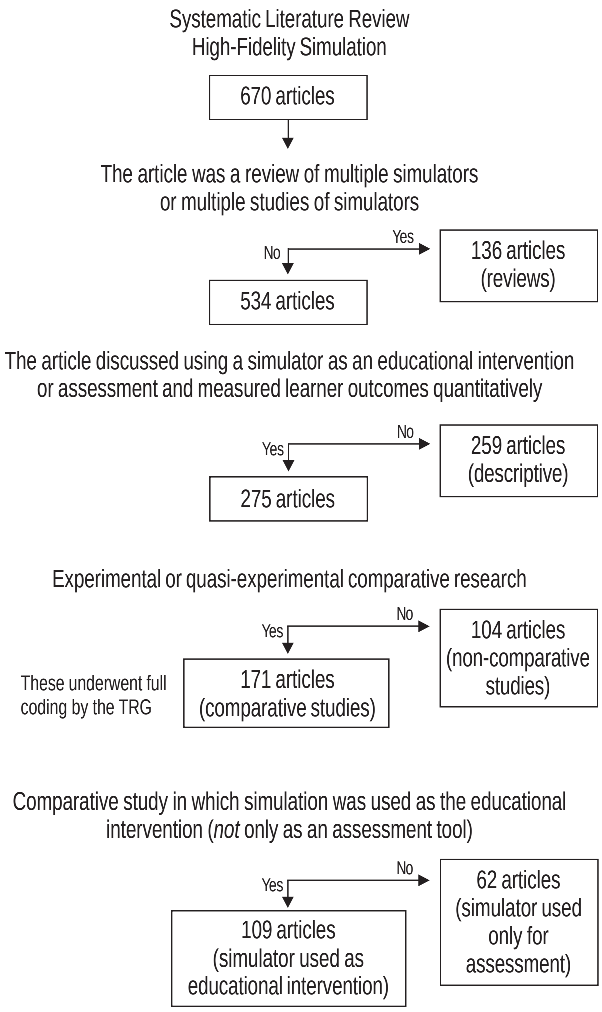


Table 2: Literature search strategy

Search Terms/Search Concepts				
1 Simulator	20 Trauma	39 Medical student	58 Large group	75 Certification
2 Simulation	21 Dental	40 Graduate	59 Lecture	76 Validity
3 Mannikin	22 Nursing	41 Resident	60 Small group	77 Reliability
4 Human model	23 Endovascular	42 Continuing education	61 Instructor	78 Feasibility
5 Virtual reality	24 Colonoscopy	43 Professional	62 Computer-based	79 Skills
6 Full body	25 Sigmoidoscopy	44 Practitioner	63 Clinical	80 Procedures
7 3-dimensional	26 Intravenous	45 Education	64 Peer	81 Management
8 Internal medicine	27 Arterial	46 Training	65 Classroom	82 Health promotion
9 Pediatric	28 Gastroenterology	47 Curriculum	66 Hospital	83 Communication
10 Surgery	29 Multimedia	48 Community	67 Ambulatory	84 Information
11 Orthopedic	30 Minimally invasive	49 Core	68 Laboratory	85 Attitudes
12 Cardiovascular	31 Suture	50 Optional	69 Clinical skills centre	86 Behavior
13 Endoscopic	32 Diagnostic	51 Elective	70 Distance learning	87 Decision-making
14 Laparoscopic	33 Ultrasound	52 Integrated	71 Assessment	88 Patient safety
15 Arthroscopic	34 Force feedback	53 Outcome-based	72 Testing	89 Medical errors
16 Sinus	35 Tactile	54 Problem-based	73 Evaluation	90 Team
17 Anesthesia	36 Haptic	55 Multi-professional	74 Grade	91 Development
18 Critical care	37 Undergraduate	56 Learning		
19 Emergency	38 Medical school	57 Independent		

combinations in the searches. Thus the essential pattern was the same for each search but adjustments were made for databases that enabled controlled vocabulary searching in addition to text word or keyword phrase searching. This approach acknowledges the role of "art" within information science, recognizing that information retrieval requires professional judgment coupled with high technology informatics (Ojala, 2002).

Step 4: *Research study selection***.** The literature search strategy yielded an initial pool of 670 peer reviewed journal articles or other documents (i.e., doctoral dissertations, academic meeting papers) that have undergone scholarly scrutiny. Four screening criteria were then used to reduce the initial pool to a focused set of studies: (a) elimination of review articles in favor of empirical studies; (b) use of a simulator as an educational assessment or intervention with learner outcomes measured quantitatively; (c) the research must be comparative, either experimental or quasi-experimental; and (d) research that involves simulation solely as an educational intervention, i.e., eliminating simulation-based assessment. Use of the four screening criteria resulted in a final set of 109 articles (16% of the initial pool) that form the basis of this systematic review (Figure 2).

Step 5: *Data extraction and coding.* Data were extracted systematically from the 109 eligible journal articles by the independent observers in the study phase, using the coding sheet presented in Appendix 1. A list of the 109 journal articles coded and analyzed in this study appears as Appendix 2. The 62 journal articles eliminated from this report because they address medical simulations only as assessment tools are listed in Appendix 3.

Step 6: *Data analysis and synthesis***.** Qualitative data synthesis and tabular presentation of research methods and outcomes were used. Heterogeneity of research designs, educational interventions, outcome measures, and timeframe precluded data synthesis using meta-analysis. This is similar to the recent systematic review of problem based learning (PBL) in medical education, where heterogeneous research methods prevented quantitative meta-analysis of PBL outcome data (Newman and the Pilot Review Group, 2003).

Results

Coding accuracy

Coding accuracy for features and qualities of the journal articles was achieved in two ways. First, coding about the *features* of the articles that are captured in the coding sheet items found in Appendix 1 was done by consensus. Each article was read and coded by at least two TRG members. These coding judgments were then discussed openly. Any initial coding disagreements were resolved by group consensus so that all decisions about features of the articles were *unanimous.*

Second, the 109 journal articles in the final set where the simulator was used as an educational intervention were also coded for *quality* by two raters. Each rater was "blind" to the coding decisions made by his/her partner. Each article was coded against four categorical items: (a) design, (b) implementation, (c) analysis, and (d) strength of findings. Each item was rated on a scale ranging from 1 = Strongly Disagree to 3 = Uncertain to 5 = Strongly Agree.

We defined coding "agreement" either as (a) no discrepancy between the two ratings of each study item, i.e., perfect agreement, or (b) two ratings within one point on each study item. Results from the coding accuracy tabulation are shown in Table 3. The rating data show that evaluations of research study quality were in very high agreement, much higher than values ranging from .25 to .35 usually found among expert ratings of manuscripts submitted for publication in scholarly journals and for quality judgments about research grant applications (Cicchetti, 1991).

Table 3: Coding Accuracy

	Percent Agreement	
Coding Items	**Perfect**	**Within 1 Point**
1. Design	45%	86%
2. Implementation	45%	91%
3. Analysis	41%	83%
4. Conclusions	35%	81%

Research study features

Selected results obtained from the consensual coding of research study features using items contained in Appendix 1 are shown in Figure 3, panels A to G.

Figure 3: Research study features

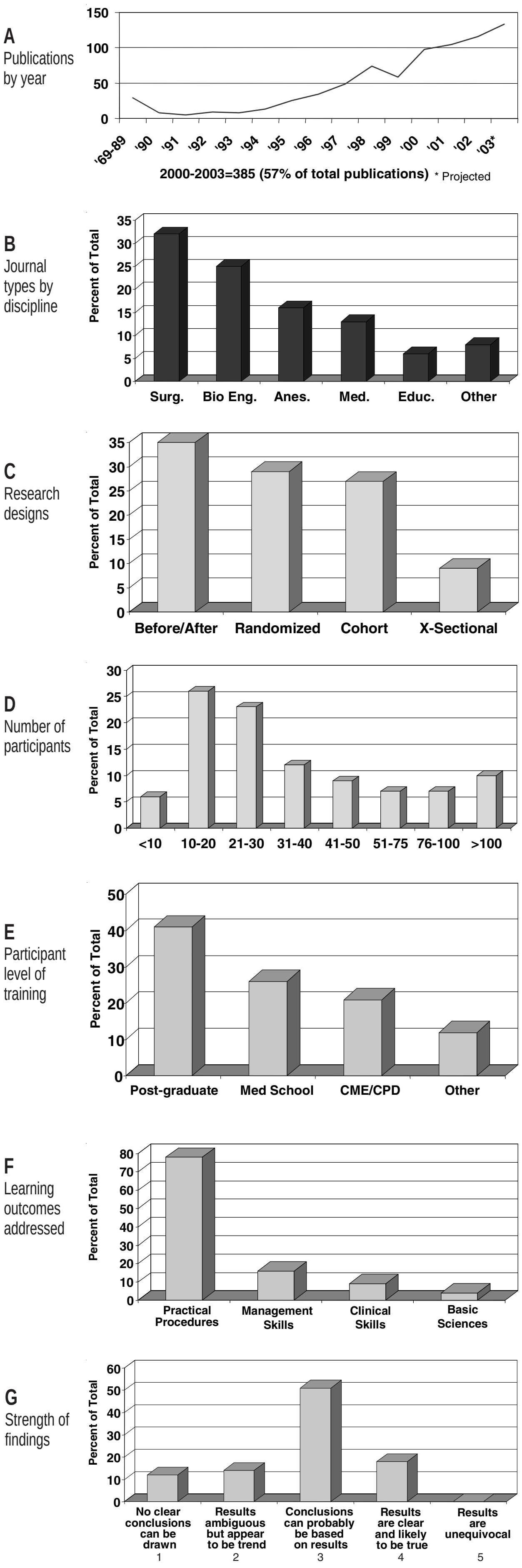


Figure 3A shows that the absolute number of journal articles about high-fidelity simulations in medical education has increased rapidly over the 34-year time span of this review. Few journal articles were published in the decades of the 1970s and 1980s. However, beginning in the early 1990s, (coincident with the availability of personal computers) the growth of high-fidelity simulation-based studies in medical education has been exponential. The brief time span from 2000 to 2003 has witnessed publication of 385 of these studies, 57% of the total.

Figure 3B documents the types of disciplinary scholarly journals that have published articles about high-fidelity simulation-based medical education. The majority of these articles (over 55%) have appeared in surgical journals and journals in biomedical engineering. Research articles have also appeared in journals addressing other disciplines including anesthesiology, internal medicine, and medical education.

The research designs represented in the journal articles we reviewed are presented in Figure 3C. The modal category, before-after studies without a control group, accounts for 35% of the total. This is followed by randomized trials, cohort studies, and cross-sectional research studies, respectively.

The number of research participants (formerly called subjects) enrolled in each of the reviewed articles is shown in Figure 3D. The majority of the published research studies are quite small – over one-half enrolled less than 30 participants.

Research participants' levels of medical training are displayed in Figure 3E. The modal research participant is a postgraduate resident in one of the medical specialties (e.g., surgery, anesthesiology). However, high-fidelity simulation journal articles have also reported research at the levels of undergraduate medical education, continuing medical education, and professional development.

Figure 3F shows clearly that journal articles reporting original research about the use of high-fidelity simulations in medical education are focused on learner acquisition of skill at performing practical procedures. Articles addressing learning outcomes in such categories as management skills, clinical skills, and knowledge of the basic medical sciences have been published with much lower frequency.

The strength of findings reported in the journal articles we reviewed is presented in Figure 3G. There is much variation in the strength of findings in these peer-reviewed publications. Approximately 80% of the reported research findings are equivocal. Less than 20% of the publications report results that are clear and likely to be true. None of the peer-reviewed journal articles report unequivocal research results as judged by our reviewers.

Simulator features, use, and effective learning

As a result of our inclusion criteria, we selected studies in which a simulator was used as an educational intervention and learner outcomes were measured, including participation, attitudes, knowledge and skills. Thus, all of the studies that were coded met one or more of Kirkpatrick's training criteria for effectiveness. Table 4 presents our qualitative distillation of the features and uses of high-fidelity medical simulations that lead to effective learning. We identified ten features and uses of the medical simulations as *educational interventions* and present them in order of the number of times they were coded (Item 10 of Appendix 1).We also include the average rating for the strength of findings for those studies associated with each feature.

1 **Feedback.** Feedback, knowledge of results of one's performance, is the single most important feature of simulation-based medical education toward the goal of effective learning. Educational feedback also appears to slow the decay of acquired skills and allows learners to self-assess and monitor their progress toward skill acquisition and maintenance. Sources of feedback may either be "built-in" to a simulator, given by an instructor in "real time" during educational sessions, or provided *post hoc* by viewing a videotape of the simulation-based educational activity. The source of the feedback is less important than its presence. Fifty-one of the 109 journal articles listed in the final stage of this review (47%) report specifically that educational *feedback to learners* is a principal feature of simulation-based medical education.

2 **Repetitive practice.** Opportunity for learners to engage in focused, repetitive practice where the intent is skill improvement, not idle play, is a basic learning feature of high-fidelity medical simulations. Repetitive practice involves intense and repetitive learner engagement in a focused, controlled domain. Skill repetition in practice sessions gives learners opportunities to correct errors, polish their performance, and make skill demonstration effortless and automatic. Outcomes of repetitive practice include skill acquisition in shorter time periods than exposure to routine ward work and transfer of skilled behavior from simulator settings to patient care settings. Of course, medical simulation devices and procedures must be time available (i.e., accommodate learner schedules) and physically convenient (i.e., close to hospital wards and clinics) so learners can practice skills repetitively. Recent research (Ericsson, 2004) underscores the importance of repetition for clinical skill acquisition and maintenance. Forty-three journal articles (39%) identified repetitive practice as a key feature involving the use of high-fidelity simulations in medical education.

3 **Curriculum integration.** Twenty-seven of the 109 studies contained in the final stage of this systematic review (25%) cite integration of simulation-based exercises into the standard medical school or postgraduate educational curriculum as an essential feature of their effective use. Simulation-based education should *not* be an *extra*-ordinary activity, but must be grounded in the ways learner performance is evaluated, and should be built into learners' normal training schedule. Effective medical learning stems from learner engagement in deliberate practice with clinical problems and devices in simulated settings in addition to patient care experience. Medical education using simulations must be a required component of the standard curriculum. Optional exercises arouse much less learner interest.

4 **Range of difficulty level.** Effective learning is enhanced when learners have opportunities to engage in practice of medical skills across a wide range of difficulty levels. Trainees begin at basic skill levels, demonstrate performance mastery against objective criteria and standards, and proceed to training at progressively higher difficulty levels. Each learner will have a different "learning curve" in terms of shape and acceleration although long-run learning outcomes, measured objectively, should be identical. Fifteen of the 109 journal articles covered in this review (14%) address the importance of the range of task difficulty level as an important variable in simulation-based medical education.

5 **Multiple learning strategies.** The adaptability of high-fidelity medical simulations to multiple learning strategies is both a feature and a use of the educational devices. This capability was identified in 11 of the 109 scientific journal articles (10%). Multiple learning strategies include but are not limited to instructor-centered education involving either (a) large groups [e.g., lectures]; or (b) small groups [e.g., tutorials]; (c) small-group independent learning without an instructor; and (d) individual, independent learning. Of course, optimal use of high-fidelity simulations in such different learning situations depends on the educational objectives being addressed and the extent of prior learning among the trainees. The rule-of-thumb is that one's educational tools should match one's educational goals. High-fidelity medical simulations that are adaptable to several learning strategies are more likely to fulfill this aim.

6 **Capture clinical variation.** High-fidelity medical simulations that can capture or represent a wide variety of patient problems or conditions are obviously more useful than simulations having a narrow patient range. Simulations capable of sampling from a broad universe of patient demographics, pathologies, and responses to

treatment can increase the number and variety of patients that learners encounter. Boosting the variety of simulated patients seen by learners helps to standardize the clinical curriculum across educational sites. This gives “equity” to smaller programs, often in remote locations, where the range of real patients may be restricted. Such simulations can also give learners exposure and practice experience with rare, life-threatening patient problems where the presentation frequency is low while the stakes are high. Eleven of the 109 journal articles (10%) cited capturing clinical variation as a key simulation feature.

7 **Controlled environment**. In a controlled clinical environment learners can make, detect, and correct patient care errors without adverse consequences, while instructors can focus on learners, not patients. High-fidelity simulations are ideal for work in controlled, forgiving environments in contrast to the uncontrolled character of most patient care settings. Education in a controlled environment allows instructors and learners to focus on “teachable moments” without distraction and take full advantage of learning opportunities. This also reflects a clinical and educational culture focused on ethical training involving learners and patients. The utility of education in a controlled environment using high-fidelity medical simulations was mentioned in 10 of the 109 journal articles (9%).

8 **Individualized learning**. The opportunity for learners to have reproducible, standardized educational experiences where they are active participants, not passive bystanders, is an important quality about the use of high-fidelity medical simulations. This means that learning experiences can be individualized for learners, adapted to one’s unique learning needs. Simulations allow complex clinical tasks to be broken down into their component parts for educational mastery in sequence at variable rates. Learners can take responsibility for their own educational progress within the limits of curriculum governance. The goal of uniform educational outcomes despite different rates of learner educational progress can be achieved with individualized learning using high-fidelity medical simulations. This feature was highlighted by 10 of the 109 journal articles (9%).

9 **Defined outcomes or benchmarks.** In addition, individualized learning in a controlled educational environment, high-fidelity medical simulations can feature clearly defined outcomes or benchmarks for learner achievement. These are plain goals with tangible, objective measures. Learners are more likely to master key skills if the outcomes are defined and appropriate for their level of training. Examples include the virtual reality metrics of Gallagher and Satava (2002) and scorecard endoscopy described by Neumann and colleagues (2003). This feature of high-fidelity medical simulations was named by seven of the 109 reviewed journal articles (6%).

10 **Simulator validity.** There are many types of educational validity both in the presentation of learning materials and events and in measuring educational outcomes. In this case, validity means the degree of realism or fidelity the simulator provides as an approximation to complex clinical situations, principles, and tasks. High simulator validity is essential to help learners increase their visiospatial perceptual skills and to sharpen their responses to critical incidents. Clinical learners prefer this realism (face validity) with opportunities for hands-on experience. Concurrent validity is frequently considered to be the generalizability of simulation-based clinical learning to real patient care settings. The issue of simulation validity was covered in four of the 109 journal articles we reviewed (3%).

Discussion

What do the findings mean?

The research evidence is clear that high-fidelity medical simulations facilitate learning among trainees when used under the right conditions. Those conditions are listed in Table 4, ranging from giving feedback to learners and providing opportunities for repetitive practice to curriculum integration, individualized learning, and simulator validity. These ten conditions represent an ideal set of educational circumstances for the use of medical simulation that can rarely be fully satisfied in all training settings. The conditions do, however, represent a set of goals for educational programs to reach to maximize the impact of simulation-based training.

The evidence also shows that simulation-based medical education complements, but does not duplicate, education involving real patients in genuine settings. Simulation-based medical education is best employed to prepare learners for real patient contact. It allows them to practice and acquire patient care skills in a controlled, safe, and forgiving environment. Skill acquisition from practice and feedback also boosts learner self-confidence and perseverance, affective educational outcomes that accompany clinical competence.

Issues including simulator cost effectiveness and incentives for product development and refinement are beyond the scope of this review. The cost effectiveness of simulation-based medical education has been addressed in many other reports (e.g., Gaba, 2000; Issenberg *et al.* 1999a, 2002) that frequently make a strong case about the costs of *not* using simulation technology in medical education. Incentives for continued development and refinement of medical simulation technology reside with entrepreneurs, chiefly in the commercial sector. These incentives will grow as research and experience demonstrate that medical education simulation works.

Limitations of the review

All scholarship has limits, rarely failures, and this review is no exception. The principal limit is that the quality and utility of the review stems directly from the quality of the primary research it covers. We reported in Figure 3G that approximately 80% of the published research findings are equivocal at best and only 20% of the research publications we reviewed report outcomes that are clear and likely true. Consequently, the state of the research enterprise in simulation-based medical education prohibits strong inference and generalizable claims about efficacy. The direction of the evidence is clear—high-technology simulations work under the right conditions.

Limits of the published body of evidence ruled-out a formal meta-analysis for this review, similar to the work

Table 4: Features and Uses of High-fidelity Simulators that Lead to Effective Learning
(Study ID refers to references listed in Appendix 2)

Features and uses	Number of studies	Strength of findings	Study ID	Comments
Feedback is provided during learning experience	51	3.5	1, 2, 6, 10, 11, 12, 13, 16, 21, 23, 24, 28, 31, 32, 35, 38, 41, 42, 46, 47, 50, 51, 52, 58, 59, 61, 62, 63, 64, 70, 71, 72, 73, 75, 78, 79, 80, 81, 87, 88, 91, 92, 93, 94, 99, 100, 101, 103, 104, 105, 107	Slows decay in skills over time; self assessment allows individual to monitor progress; can be 'built-in' to simulator or provided by instructor immediately or later via video-taped debriefing
Learners engage in repetitive practice	43	3.2	1, 2, 5, 12, 16, 19, 26, 28, 32, 33, 34, 38, 39, 40, 41, 42, 43, 46, 47, 50, 51, 53, 54, 55, 59, 69, 70, 73, 75, 80, 81, 83, 86, 90, 91, 92, 94, 97, 98, 101, 105, 106, 108	Primary factor in studies showing skills transferring to real patients; shortens learning curves and leads to faster automaticity; simulator must be made available - convenient location, accommodates learner schedule
Simulator is integrated into overall curriculum	27	3.2	4, 14, 15, 16, 19, 21, 22, 24, 30, 31, 37, 39, 41, 44, 52, 56, 57, 61, 62, 63, 64, 67, 72, 75, 88, 93, 95	Simulator fully integrated into overall curriculum, eg ACLS, ATLS, CRM, basic surgical training
Learners practice with increasing levels of difficulty	15	3.0	7, 17, 22, 28, 32, 33, 34, 35, 47, 48, 51, 54, 73, 99, 100	Increasing degree of difficulty increases mastery of skill
Adaptable to multiple learning strategies	11	3.2	21, 24, 25, 26, 39, 44, 46, 72, 74, 95, 107	Simulator used instructor large-group and small-group settings; independent small-group and individual settings
Clinical Variation	11	3.1	4, 9, 20, 26, 27, 81, 84, 95, 96, 99, 100	Can increase the number and variety of patients a learner encounters; provides equity to smaller training programs; provides exposure to rate encounter
Controlled Environment	10	3.2	2, 19, 20, 26, 46, 75, 82, 85, 95, 96	Learners make and detect mistakes without consequences; instructors can focus on learners through 'teachable moments'; reflects educational 'culture' focused on ethical training
Individualized Learning	10	3.3	1, 16, 21, 26, 31, 46, 52, 72, 88, 109	Provides reproducible, standardized experience for all learners; learner is active participant, responsible for his/her own learning
Outcomes/benchmarks clearly defined	7	3.1	1, 29, 31, 62, 63, 64, 90	Learners more likely to master skill if outcomes are clearly defined and appropriate for learner level of training
Validity of simulator	4	2.9	8, 18, 22, 99	Face validity - realism provides context for understanding complex principles/tasks, increases visiospatial perceptual skills, learners prefer realism; concurrent validity - ability on simulator transfers to real patient

of Newman (2003) who attempted a meta-analysis of research on problem-based learning. Heterogeneity of research designs and study quality, unstandardized outcome measures, and wide variation in details given in journal articles (e.g., many fail to report means, standard deviations, and reliability coefficients) make a quantitative synthesis of the research evidence impossible.

Research agenda

The lack of unequivocal evidence for much of the research on simulation-based medical education clearly calls for better research and scholarship in this sector of medical education. Responsibility resides not only with investigators who plan and execute research studies but also with journal editors and editorial boards who evaluate submitted manuscripts and set quality standards. Studies that feature weak designs, small samples, inattention to psychometric properties of variables, and flawed analyses lack rigor and do not advance educational science. Journal articles that lack details about data and methods prevent clear interpretation and prevent replication. As pointed out by Colliver (2003), Lurie (2003), and the Joint Task Force of *Academic Medicine* and the GEA-RIME Committee (2001) medical education research needs much improvement to advance knowledge and inform practice. An additional outcome of this BEME project was the development of more formal guidelines for those who wish to carry out educational studies involving simulators (Figure 4).

An untouched research area that is suited perfectly to high-fidelity simulation in medical education concerns the introduction of mastery learning models. In brief, mastery learning aims to produce identical outcomes for all at high performance standards. Time needed to achieve mastery is the variable in the educational equation. For example, if the educational goal is cardiac auscultation at 90% accuracy, then medical learners are allowed to practice deliberately with a cardiac patient simulator for the time needed to achieve the standard. In mastery learning, outcomes are uniform while the time needed to reach them varies (Bloom, 1974, 1976; Carroll, 1963). Mastery learning is also a key component of competency-based education (McGaghie *et al.* 1978).

Qualitative studies also have a place on the high-fidelity research agenda in medical education. We need to know more about how to establish and maintain a positive and energetic learning atmosphere in medical simulation centers. This will encourage medical learners at all levels to seek simulation-based education because it will help them become superb clinicians. The moment a medical simulation center is perceived to be a "shooting gallery," focused on learner problems and deficiencies, not improvement, its educational effectiveness is ruined. This acknowledges the widespread phobia of evaluation apprehension among medical learners (DelVecchio Good, 1995; McGaghie *et al*, 2004) and the need to reduce its influence.

An additional observation from this study warrants mention. We noted in retrospect, but did not code prospectively, that few research studies in each of the clinical medical specialties cite research *outside* of their own field. Anesthesiologists cite the anesthesiology literature, surgeons highlight studies reported in surgical journals, computer specialists and technocrats look inward. Few high-fidelity medical simulation journal articles cite the general medical education literature, much less articles in business and industry, aviation, and the military. There appears to be little awareness of the substantive and methodological breadth and depth of educational science in this field. We conclude that investigators need to be *better informed* if simulation-based medical education is to advance as a discipline.

Figure 4: Guidelines for educational studies involving simulators

Appropriateness of Study Design
1. Clear statement of the research question
2. Awareness of current state-of-affairs (literature)
3. Clear specification of:
   a. population
   b. sample from population
4. Intervention description
   a. frequency
   b. duration
5. Prospective vs. retrospective
6. Random selection of subjects vs. non-random sampling
7. Evidence for pre-study equivalence of groups
8. Is the outcome measure the proper one for the study?
9. Report of measurement characteristics of outcomes
   a. reliability
   b. validity
10. Pre-intervention measurement: Yes/No
11. Follow-up outcome measurement (maintenance of effect): Yes/No

Implementation of Study Adequate
12. Little or no attrition vs. more attrition: how much?
13. Simulator characteristics
    a. reliability (consistent operation of simulation)
    b. validity (e.g., differentiates novice and experts)

Appropriate Data Analysis
14. Correct analytic approaches: Yes/No
15. Larger effect size vs. smaller effect size
16. Statistical significance: Yes/No
17. Practical performance standard specified: Yes/No
18. Results meet or exceed performance standard: Yes/No
19. Evidence that results generalize to clinical practice: Yes/No

Quality of Conclusions & Recommendations
20. Conclusions and recommendations supported and consistent with size of results

Conclusions

This **BEME report** is the first systematic review of the research evidence about the features and use of high-fidelity medical simulations that lead to effective learning. Our goal was to cover the scientific literature comprehensively, with detail and rigor. The intent was to paint an objective portrait of the current state of knowledge about high-fidelity simulation in medical education and to begin to set an agenda for continued evaluation research. We are successful to the degree that readers are better informed about this medical education innovation and are motivated to advance simulation-based medical education via advocacy, teaching, and research.

The **report** began with a broad and deep **introduction** to the 34-year history and present use of high-fidelity simulation in medical education. The approach we used to conduct the systematic review is described in detail in the **methods** section. Our **results** are presented in three parts: (a) coding accuracy; (b) research study features; and (c) simulator features, use, and effective learning. In our discussion section we present our **conclusions** in three categories: (a) What do the findings mean? (b) limitations of the review; and (c) research agenda.

Our goal in this project was to determine from the existing literature the best evidence for using high-fidelity simulation in medical education. We did not evaluate whether simulators are more or less effective than traditional or alternative methods. We would have very likely come to the same conclusions as others when comparing one type of educational intervention with another (Newman & the Pilot Review Group, 2003; Dolmans, 2003). Instead, we purposely selected articles that demonstrated effective learning at least at the level of participation and, in most cases, an improvement in knowledge, skills and attitudes. This enabled us to review and evaluate the existing evidence, and to distill several important features and aspects of simulators that will lead to effective learning:

- Provide feedback during the learning experience with the simulator
- Learners should repetitively practice skills on the simulator
- Integrate simulators into the overall curriculum
- Learners should practice with increasing levels of difficulty (if available)
- Adapt the simulator to complement multiple learning strategies
- Ensure the simulator provides for clinical variation (if available)
- Learning on the simulator should occur in a controlled environment
- Provide individualized (in addition to team) learning on the simulator
- Clearly define outcomes and benchmarks for the learners to achieve using the simulator
- Ensure the simulator is a valid learning tool.

References

ABRAHAMSON, S., DENSON, J.S. & WOLF, R.M. (1969) Effectiveness of a simulator in training anesthesiology residents, *Journal of Medical Education*, 44, pp. 515-519.

ACGME OUTCOMES PROJECT. Accreditation Council for Graduate Medical Education website. Available at http://www.acgme.org.2000. Accessed 2 August 2003.

AVIATION WEEK & SPACE TECHNOLOGY (1997). Simulator trained Bush for a voluntary jump. (28 April), 146, p. 62.

BARACH, P. & MOSS, F. (2002) Delivering safe health care: safety is a patient's right and the obligation of all health professionals, *Quality Health Care*, 10, pp. 199-203.

BLOOM, B.S. (1974) Time and learning. *American Psychologist*, 29, pp. 682-688.

BLOOM, B.S. (1976) *Human Characteristics and School Learning* (New York, McGraw-Hill).

BOGNER, M.S. (Ed.). (1994) *Human Error in Medicine* (Hillsdale, N.J., Lawrence Erlbaum Associates).

BRANNICK, M.T., SALAS, E. & PRINCE, C. (1997) *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (Mahwah, N.J., Lawrence Erlbaum Associates).

BRENNAN, T.A., LEAPE, L.L., LAIRD, N.M., HEBERT, L., LOCALIO, A.R., LAWTHERS, A.E., NEWHOUSE, J.P., WEILER, P.C. & HIATT, H.H. (1991) Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study. *New England Journal of Medicine*, 324, pp. 370-376.

CARROLL, J.B. (1963) A model of school learning. *Teachers College Record*, 64, pp. 723-733.

CICCHETTI, D.V. (1991) The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14, pp. 119-186.

COLLIVER, J.A. (2003) The research enterprise in medical education, *Teaching and Learning in Medicine*, 15, pp. 154-155.

DEL VECCHIO GOOD, M.J. (1995) *American Medicine: The Quest for Competence.* (Berkeley, California, University of California Press).

DEZIEL, D.J., MILLIKAN, K.W., ECONOMOU, S.G., DOOLAS, A., KO, S.T. & AIRAN, M.C. (1993) Complications of laparoscopic cholecystectomy: a national survey of 4,292 hospitals and an analysis of 77,604 cases. *American Journal of Surgery*, 165, pp. 9-14.

DOLMANS, D. (2003) The effectiveness of PBL: the debate continues. Is meta-analysis helpful? *Medical Education*, 37, pp. 1129-1130.

ERICSSON, K.A. (2004) Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, 79 (10, Suppl.), pp. 570-581.

ERICSSON, K.A. & CHARNESS, N. (1994) Expert performance: its structure and acquisition, *American Psychologist*, 49, pp. 725-747.

ERICSSON, K.A. & LEHMANN, A.C. (1996) Expert and exceptional performance: evidence of maximal adaptation to task constraints, *Annual Review of Psychology*, 47, pp. 273-305.

ERICSSON, K.A., KRAMPE, R.T & TESCH-RÖMER, C. (1993) The role of deliberate practice in the acquisition of expert performance, *Psychological Review*, 100, pp. 363-406.

EYLER, A.E., DICKEN, L.L., FITZGERALD, J.T., OH, M.S., WOLF, F.M. & ZWEIFLER, A.J. (1997) Teaching smoking-cessation counseling to medical students using simulated patients. *American Journal of Preventive Medicine*, 13, pp. 153-158.

FINCHER, R.M.E. & LEWIS L.A. (2002) Simulations used to teach clinical skills, in: G.R. Norman, C.P.M.van der Vleuten & D.I. Newble (Eds.) International *Handbook of Research in Medical Education, Part One (*Dordrecht, The Netherlands, Kluwer Academic Publishers).

GABA, D. (2000) Human work environment and simulators, in: R.D. Miller (Ed.) Anesthesia, 5th ed. (Philadelphia, Churchill Livingstone).

GALLAGHER, K.A. & SATAVA, R.M. (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. *Surgical Endoscopy*, 16(12), pp. 1746-1752.

HALUCK, R.S., MARSHALL, R.L., KRUMMEL, T.M. & MELKONIAN, M.G. (2001) Are surgery training programs ready for virtual reality? A survey of program directors in general surgery, *Journal of the American College of Surgeons*, 193, pp. 660-665.

HARDEN, R.M., GRANT, J., BUCKLEY E.G. & HART, I.R. (1999) BEME guide no. 1: Best Evidence Medical Education, *Medical Teacher*, 21, pp. 553-562.

HELMREICH, R.L. & SCHAEFER, H-G. (1994) Team performance in the operating room, in: M.S. Bogner (Ed.) *Human Error in Medicine (*Hillsdale, N.J., Lawrence Erlbaum Associates)

ISSENBERG, S.B., GORDON, M.S., GORDON, D.L., SAFFORD, R.E. & HART, I.R. (2001) Simulation and new learning technologies, *Medical Teacher*, 16, pp. 16-23.

ISSENBERG, S.B., MCGAGHIE, W.C., BROWN, D.D., MAYER, J.W., GESSNER, I.H., HART, I.R., WAUGH, R.A., PETRUSA, E.R., SAFFORD, R., EWY, G.A. & FELNER, J.M. (2000) Development of multimedia computer-based measures of clinical skills in bedside cardiology, in: D.E. MELNICK (Ed.) The *Eighth International Ottawa Conference on Medical Education and Assessment Proceedings. Evolving Assessment:Protecting the Human Dimension* (Philadelphia, National Board of Medical Examiners).

ISSENBERG, S.B., MCGAGHIE, W.C., GORDON, D.L., SYMES, S., PETRUSA, E.R., HART, I.R. & HARDEN, R.M. (2002) Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice, *Teaching and Learning in Medicine*, 14, pp. 223-228.

ISSENBERG, S.B., MCGAGHIE, W.C., HART, I.R., MAYER, J.W., FELNER, J.M., PETRUSA, E.R., WAUGH, RA, BROWN, D.D., SAFFORD, R.R., GESSNER, I.H., GORDON, D.L. & EWY, G.A. (1999a) Simulation technology for health care professional skills training and assessment. *Journal of the American Medical Association*, 282, pp. 861-866.

ISSENBERG, S.B., PETRUSA, E.R., MCGAGHIE, W.C., FELNER, J.M., WAUGH, R.A., NASH, I.S. & HART, I.R. (1999b) Effectiveness of a computer-based system to teach bedside cardiology, *Academic Medicine*, 74 (10, Suppl.), pp. S93-S95.

JOINT TASK FORCE OF *ACADEMIC MEDICINE* AND THE GEA-RIME COMMITTEE (2001) Review criteria for research manuscripts, *Academic Medicine*, 76, pp. 898-978.

KIRKPATRICK, D.I. (1998) *Evaluating Training Programs: The Four Levels*, 2nd ed. (San Francisco, Berrett-Koehler Publishers).

KOHN, L., CORRIGAN, J. & DONALDSON, M. (1999) *To Err is Human: Building a Safer Health System (*Washington, DC, National Academy Press).

LURIE, S.J. (2003) Raising the passing grade for studies of medical education, *Journal of the American Medical Association*, 290, pp. 1210-1212.

MANGIONE, S. & NIEMAN, L.Z. (1997) Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency, Journal *of the American Medical Association*, 278, pp. 717-722.

MCGAGHIE, W.C., MILLER, G.E., SAJID, A. & TELDER, T.V. (1978) *Competency-Based Curriculum Development in Medical Education. Public Health Paper No. 68.* (Geneva, Switzerland, World Health Organization).

MCGAGHIE WC. (1999) Simulation in professional competence assessment: basic considerations, in: A. Tekian, C.H. McGuire & W.C. McGaghie (Eds) *Innovative Simulations for Assessing Professional Competence* (Chicago, Department of Medical Education, University of Illinois at Chicago).

MCGAGHIE, W.C., DOWNING, S.M. & KUBILIUS, R. (2004) What is the impact of commercial test preparation courses on medical examination performance? *Teaching and Learning in Medicine*, 16, pp. 202-211.

MELLER, G. (1997) A typology of simulators for medical education. *Journal of Digital Imaging*, 10 (3, Suppl. 1, August), pp. 194-196.

MILLER, G.E. (1990) The assessment of clinical skills/ competence/performance, *Academic Medicine*, 65 (Suppl. 9), pp. S63-S67.

MILLOS, R.T., GORDON, D.L., ISSENBERG, S.B., REYNOLDS, P.S., LEWIS, S.L., MCGAGHIE, W.C., PETRUSA, E.R. & GORDON, M.S. (2003). Development of a reliable multimedia computer-based measure of clinical skills in bedside neurology. *Academic Medicine* 78 (10, Suppl), pp. S52-S54.

NEUMANN, M., SIEBERT, T., RAUSCH, J., HORBACH, T., ELL, C., MANEGOLD, C., HOHENBERGER, W. & SCHNEIDER, I. (2003). Scorecard endoscopy: a pilot study to assess basic skills in trainees for upper gastrointestinal endoscopy. *Langenbecks Archives of Surgery* 387 (9-10), pp. 386-391.

NEWMAN, M. and the Pilot Review Group. (2003) A pilot systematic review and meta-analysis on the effectiveness of problem based learning. *Newcastle: Learning and Teaching Subject Network for Medicine, Dentistry ad Veterinary Medicine.* http://www.ltsn01.ac.uk/resources/features/pbl.

OJALA, M. (2002) Information professionals as technologists, *Online*, 26, p. 5.

PUGH, C.M. & YOUNGBLOOD, P. (2002) Development and validation of assessment measures for a newly developed physical examination simulator, *Journal of the American Medical Informatics Association*, 9, pp. 448-460.

REEVES, S., KOPPEL, I., BARR, H., FREETH, D. & HAMMICK, M. (2002) Twelve tips for undertaking a systematic review, *Medical Teacher*, 24, pp. 358-363.

ROLDAN, C.A., SHIVLEY, B.K. & CRAWFORD, M.H. (1996) Value of the cardiovascular physical examination for detecting valvular heart disease in asymptomatic subjects, *American Journal of Cardiology*, 77, pp. 1327-1331.

ROLFE, J.M. & STAPLES, K.J. (1986) *Flight Simulation (*Cambridge, Cambridge University Press).

SCHAEFER, J.J., DONGILLI, T., & GONZALEZ, R.M. (1998) Results of systematic psychomotor difficult airway training of residents using the ASA difficult airway algorithm & dynamic simulation, *Anesthesiology,* 89(3A), Supplement, A60.

SELIGMAN J. (1997, 7 April) Presidential high: more than 50 years after a tragic wartime jump, George Bush has a happier landing. *Newsweek*, 129, p. 68.

SPENCER, L.M., SPENCER, S.M. (1993) *Competence at Work: Models for Superior Performance (*New York, John Wiley & Sons).

TEKIAN, A., MCGUIRE, C.G. & MCGAGHIE, W.C. (Eds.). (1999) *Innovative Simulations for Assessing Professional Competence (*Chicago, Department of Medical Education, University of Illinois at Chicago).

THORNTON, G.C., MUELLER-HANSON, R.A. (2004) *Developing Organizational Simulations: A Guide for Practitioners and Students.* (Mahwah, NJ, Lawrence Erlbaum Associates).

WAYNE, D.B., BUTTER, J., SIDDALL, V., FUDALA, M., LINDQUIST, L., FEINGLASS, J., WADE, L.D., & MCGAGHIE, W.C. (2005) Simulation-based training of internal medicine residents in advanced cardiac life support protocols: a randomized trial, *Teaching and Learning in Medicine*, 17, in press.

WILLIAMS, R.G., KLAMEN, D.A. & MCGAGHIE, W.C. (2003) Cognitive, social and environmental sources of bias in clinical competence ratings, *Teaching and Learning in Medicine*, 15, pp. 270-292.

WOEHR, D.J. & HUFFCUTT, A.I. (1994) Rater training for performance appraisal: a quantitative review, *Journal of Occupational and Organizational Psychology*, 67, pp. 189-205.

WOLF, F.M. (2000) Lessons to be learned from evidence-based medicine: practice and promise of evidence-based education, *Medical Teacher*, 22, pp. 251-259.

ZAHN, C. & MILLER, M.R. (2003) Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization, *Journal of the American Medical Association*, 290, pp. 1868-1874.

Funding/support

This project was supported by the State of Florida Department of Education; the Friends for Life Volunteer Organization, Miami, FL; the Ethel and W. George Kennedy Family Foundation, Coral Gables, FL; the Hugoton Foundation, New York, NY; the Madeline and Bernard Sternlight Estate, Miami, FL; and the Shepherd Broad Foundation, Inc., Miami, FL.

Acknowledgements

Michael S. Gordon, MD, PhD, Ronald M. Harden, MD, Ian R. Hart, MD, Pat Lilley, BA, and Alex Haig, MA. provided administrative or intellectual contributions to the project. We also acknowledge the administrative and technical staff of the Center for Research in Medical Education of the University of Miami School of Medicine for their continuous support of our research and scholarship.

Appendix 1 - BEME Coding Sheet

1 Administrative

Reference Number: ____________________ Reviewer: ____

Type:

☐ Book	☐ Guidelines	☐ Letter	☐ Report
☐ Comment	☐ Interview	☐ News	☐ Statistical data
☐ Conf. Paper/Proceedings	☐ Journal article	☐ Non-peer review article	☐ Test(s)
☐ Editorial	☐ Lecture	☐ Offical publication	☐ Thesis

Citation Information:

Authors: ____

Title: ____

Publication: ________________________________ Vol;Issue;Pages ________________

Search Method:

☐ Electronic search ☐ Hand search ☐ Grey literature ☐ Recommendation

2 Evaluation Methods

a Research design (tick all that apply)

Non-comparative studies

	Implied	Stated		Implied	Stated		Implied	Stated
Audit	☐	☐	Expert opinion	☐	☐	Narrative	☐	☐
Action-based	☐	☐	Focus group	☐	☐	Observation	☐	☐
Case series	☐	☐	Historical	☐	☐	Survey	☐	☐

If any of the above categories were checked, further define the approach used:

Cites evidence w/data	☐	Conceptual	☐
Descriptive	☐	Commentary	☐

Comparative studies

	Implied	Stated		Implied	Stated
Cross sectional	☐	☐	Case control	☐	☐

Single group studies	Implied	Stated	Cohort study	Implied	Stated	Trials	Implied	Stated
Before & after studies	☐	☐	Prospective	☐	☐	Non-randomized	☐	☐
Time series	☐	☐	Retrospective	☐	☐	Randomized	☐	☐

	Implied	Stated
Review	☐	☐

If any of the above categories were checked, further define the approach used:

Cites evidence w/data	☐	Conceptual	☐
Descriptive	☐	Commentary	☐

Meta-analysis ☐

b Data collection methods

☐ Interview	☐ Observation	☐ Opinion	☐ MCQ exam
☐ Patient outcomes	☐ Questionnaire	☐ Data from simulator/simulation	

3 Expected learning outcomes of intervention or approach (tick all that apply)

This section relates to the *intended* or *expected* learning outcomes for the educational intervention or educational approach described. This is different from the *impact* of the study (section 6).

	Implied	Stated		Implied	Stated
Clinical Skills	☐	☐	Understanding basic/clinical sciences	☐	☐
Practical procedures	☐	☐	Appropriate attitudes	☐	☐
Patient investigation	☐	☐	Appropriate decision making	☐	☐
Patient management	☐	☐	Role of health professional	☐	☐
Health promotion	☐	☐	Personal development	☐	☐
Communication	☐	☐	Patient safety/reducing errors	☐	☐
Appropriate information skills	☐	☐			

Please indicate specific skill (i.e., cardiac exam, intubation, venipuncture, etc) ____________________

____

4 Context (Target Population), if applicable

Number of Subjects / Size of Group ____________________

Country / Location of study ____________________

Duration of exposure: ☐ <10 hours ☐ 10 – 100 hours ☐ 100+ hours ☐ Other

Frequency of exposure: ☐ <2X ☐ 2-5X ☐ 6-10X ☐ >10X
(# of times/episodes simulator was used)

Level / stage:

☐ Pre-college	☐ CPD/CME
☐ Undergraduate college	☐ Higher general other than healthcare professional
☐ Medical school	☐ Undergraduate other than healthcare professional (early)
☐ Residency	☐ Undergraduate other than healthcare professional (later)
☐ Fellowship	☐ Pre-registration/basic professional education

Profession:

☐ Health sciences ☐ Non-health sciences
☐ Primary care medicine (internal medicine, family medicine, pediatrics, Ob/Gyn)
☐ Medicine sub-specialty (cardiology, GI, pulm)
☐ Surgery
☐ Anesthesia
☐ Dentistry
☐ Nursing
☐ Professions allied to medicine
☐ Veterinary

Please indicate specific skill (i.e., cardiology, emergency medicine, anesthesiology etc) ____________________

____

Certification:

☐ High stake (licensing, board certification), requirement for promotion to next level of training
☐ Low stake (requirement to pass course, components for all grade/evaluation, etc)
☐ Unclear / not applicable

5 Stated aim of study

Aim / objective of item ☐ Implied ☐ Stated

____

____

____

6 Impact of intervention studied

Code the level of impact being studied in the item and summarize any results of the intervention at the appropriate level. Note: include both predetermined and unintended outcomes.

Kirkpatrick hierarchy

Level 1 **Participation** - covers learners' views on the learning experience, its organization, presentation, content, teaching methods, and aspects of the instructional organization, materials, quality of instruction

____

____

Level 2a **Modification of attitudes/perceptions** - outcomes here relate to changes in the reciprocal attitudes or perceptions between participant groups toward intervention/simulation

____

____

Level 2b **Modification of knowledge/skills** - for *knowledge*, this relates to the acquisition of concepts, procedures and principles; for *skills* this relates to the acquisition of thinking/problem-solving, psychomotor and social skills

____

____

Level 3 **Behavioral change** - documents the transfer of learning to the workplace or willingness of learners to apply new knowledge & skills.

____

____

Level 4a **Change in organizational practice** - wider changes in the organizational delivery of care, attributable to an educational program

____

____

Level 4b **Benefits to patient / clients** - any improvement in the health & well being of patients/clients as a direct result of an educational program.

____

____

7 Rate evaluation methods

	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree
1 Appropriateness of study/review design	☐	☐	☐	☐	☐
2 Implementation of study/review design	☐	☐	☐	☐	☐
3 Appropriateness of data analysis	☐	☐	☐	☐	☐

4 Comment on evaluation methods, if applicable: ____

____

8 Strength of findings

	Low 1	2	3	4	High 5
1 No clear conclusions can be drawn. Not significant	☐	☐	☐	☐	☐
2 Results ambiguous, but there appears to be a trend	☐	☐	☐	☐	☐
3 Conclusions can probably be based on the results	☐	☐	☐	☐	☐
4 Results are clear and very likely to be true	☐	☐	☐	☐	☐
5 Results are unequivocal	☐	☐	☐	☐	☐

9 Educational descriptors (tick all that apply)

To provide a conceptual context of study, please check the educational descriptors for each of the following categories:

Curriculum:	☐ Community oriented	☐ Integrated, horizontal	☐ Multiprofessional	☐ Outcome-based
	☐ Core	☐ Integrated, vertical	☐ Options/special study module	☐ Problem-based
Learning:	☐ Clinical	☐ Co-operative	☐ Lectures	☐ Small group
	☐ Computer	☐ Independent	☐ Simulation	☐ Work-based
Physical setting:	☐ Classroom	☐ Ambulatory care	☐ Training center	☐ Distance learning
	☐ Teaching hospital	☐ Laboratory	☐ Clinical experience	☐ Other ___________
Assessment:	☐ Feedback	☐ Portfolio	☐ Practical	☐ Written

10 Educational features and uses of simulation (tick all that apply)

	Implied	Stated
Driven by valid curriculum-based educational need	☐	☐
Integrated into curriculum	☐	☐
Outcomes clearly defined to learner	☐	☐
Authenticity and realism (validity) of the simulation	☐	☐
Ability to control learning environment	☐	☐
User-friendly for learner	☐	☐
User-friendly for instructor	☐	☐
The presence of feedback	☐	☐
Allows independent learning	☐	☐
Opportunity to change level of difficulty	☐	☐
Allows instructor-based learning	☐	☐
High/low degree of maintenance	☐	☐
Need for support of faculty and other personnel	☐	☐
A method of documenting learner performance	☐	☐
Use at multiple learner levels	☐	☐
Use for multiple health-professional categories	☐	☐
Ability to provide variety of clinical conditions	☐	☐
Suitability for individual	☐	☐
Suitability for small-group	☐	☐
Suitability for large-group learning	☐	☐
Role of teacher clearly defined	☐	☐
Document participation	☐	☐
Document level of skill	☐	☐
Opportunity for repetitive practice	☐	☐

Additional feature/use (licensure/certification,etc) ____

11 Specific features of simulation

Indicate type / name of simulator ______________________________

Indicate manufacturer / company of simulator ______________________________

Specific capabilities of simulation / simulator (tick all that apply)

- Simulates bedside findings (heart sounds, respiration) ☐
- Simulates procedure(endoscopy, intubation) ☐
- Responds and reacts to user (anesthesia induction, ACLS) ☐
- Other imbedded simulation capability ______________________________

12 Documented improvement in learning/performance

If the article provided documentation for improvement in learning / performance, check the appropriate area:

Cognitive knowledge

Skills (demonstrated in): ☐ Simulation ☐ Real patient environment

- ☐ Hands-on psychomotor skills
- ☐ Management decision skills
- ☐ High-level communication skills

Attitude, where appropriate

Other ______________________________

13 Overall impression of article

Please make any additional comments regarding the overall *strengths* and *weaknesses* of the article.

____

____

____

____

____

____

Appendix 2 – Comparative studies with simulator used as educational intervention

1 Abrahamson, S., Denson, J.S., & Wolf R.M. (1969) Effectiveness of a simulator in training anesthesiology residents, *Journal of Medical Education*, 44(6), pp. 515-519.

2 Agazio, J.B., Pavlides, C.C., Lasome, C.E., Flaherty, N.J., & Torrance, R.J. (2002) Evaluation of a virtual reality simulator in sustainment training. *Military Medicine*, 167(11), pp. 893-897.

3 Ahlberg, G., Heikkinen, T., Iselius, L., Leijonmarck, C.E., Rutqvist, J., & Arvidsson, D. Does training in a virtual reality simulator improve surgical performance? *Surgical Endoscopy,*16, pp. 126-129.

4 Ali, J., Gana, T.J., & Howard, M. (2000) Trauma mannequin assessment of management skills of surgical residents after advanced trauma life support training. *Journal of Surgical Research*, 93(1), pp. 197-200.

5 Ali, M.R., Mowery, Y., Kaplan, B., & DeMaria, E.J. (2002) Training the novice in laparoscopy. More challenge is better. *Surgical Endoscopy*, 16(12), pp. 1732-1736.

6 Ashurst, N., Rout, C.C., Rocke, D.A., & Gouws, E. (1996) Use of a mechanical simulator for training in applying cricoid pressure. *British Journal of Anaesthesia*, 77(4), pp. 468-472.

7 Begamaschi, R., & Dicko, A. (2000) Instruction versus passive observation: a randomized educational research study on laparoscopic suture skills. *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, 10(5), pp. 319-322.

8 Brehmer, M., & Tolley, D. (2002) Validation of a bench model for endoscopic surgery in the upper urinary tract. *European Urology*, 42(2), pp. 175-179, discussion p. 180.

9 Burdea, G., Patounakis, G., Popescu, V., & Weiss, R.E. (1999) Virtual reality-based training for the diagnosis of prostate cancer. *IEEE Transactions on Biomedical Engineering*, 46(10), pp. 1253-1260.

10 Byrne, A.J., Sellen, A.J., Jones, J.G., Aitkenhead, A.R., Hussain, S., Gilder, F., Smith, H.L., & Ribes, P. (2002) Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: a multicentre study. *Anaesthesia,* 57(2), pp. 176-179.

11 Chang, K.K., Chung, J.W., & Wong, T.K. (2002) Learning intravenous cannulation: a comparison of the conventional method and the CathSim Intravenous Training System. *Journal of Clinical Nursing*, 11(1), pp. 73-78.

12 Chaudhry, A., Sutton, C., Wood, J., Stone, R., & McCloy, R. (1999) Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. *Annals of the Royal College of Surgeons of England*, 81, pp. 281-286.

13 Chopra, V., Gesink, B.J., de Jong, J., Bovill, J.G., Spierdijk, J., & Brand, R. (1994) Does training on an anaesthesia simulator lead to improvement in performance? *British Journal of Anaesthesia*, 73(3), pp. 293-297.

14 Chung, J.Y., & Sackier, J.M. (1998) A method of objectively evaluating improvements in laparoscopic skills. *Surgical Endoscopy*, 12, pp. 1111-1116.

15 Clancy, J.M., Lindquist, T.J., Palik, J.F., & Johnson, L.A. (2002) A comparison of student performance in a simulation clinic and a traditional laboratory environment: three-year results. *Journal of Dental Education*, 66(12), pp. 1331-1337.

16 Colt, H.G., Crawford, S.W., & Galbraith, O. 3rd. (2001) Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest*, 120(4), pp. 1333-1339.

17 Crossan, A., Brewster, S., Reid, S., & Mellor, D. (2002) Comparison of simulated ovary training over six different skill levels. In Proceedings of Eurohaptics 2001 (Birmingham, UK), pp. 17-21. [www.dcs.gla.ac.uk/~stephen/papers/Eurohaptics2001_crossan.pdf, accessed 20 August 2003].

18 Derosis, A.M., Fried, G.M., Abrahamowicz, M., Sigman, H.H., Barkun, J.S., & Meakins, J.L. (1998) Development of a model for training and evaluation of laparoscopic skills. *The American Journal of Surgery*, 175, pp. 482-487.

19 Derosis, A.M., Bothwell, J., Sigman, H.H., & Fried, G.M. (1998) The effect of practice on performance in a laparoscopic simulator. *Surgical Endoscopy*, 12, pp. 1117-1120.

20 Dobson, H.D., Pearl, R.K., Orsay, C.P., Rasmussen, M., Evenhouse, R., Ai, Z., Blew, G., Dech, F., Edison, M.I., Silverstein, J.C., & Abcarian H. (2003) Virtual reality: new method of teaching anorectal and pelvic floor anatomy. *Diseases of the Colon & Rectum*, 46(3), pp. 349-352.

21 Done, M.L., & Parr, M. (2002) Teaching basic life support skills using self-directed learning, a self-instructional video, access to practice manikins and learning in pairs. *Resuscitation*, 52(3), pp. 287-291.

22 Edmond, C.V. Jr. (2002) Impact of the endoscopic sinus surgical simulator on operating room performance. *Laryngoscope*, 112(7 Pt 1), pp. 1148-1158.

23 Engum, S.A., Jeffries, P., & Fisher, L. (2003) Intravenous catheter training system: computer-based education versus traditional learning methods. *The American Journal of Surgery*, 186(1), pp. 67-74.

24 Euliano, T.Y. (2001) Small group teaching: clinical correlation with a human patient simulator. *Advances in Physiology Education*, 25(1-4), pp. 36-43.

25 Euliano, TY. (2000) Teaching respiratory physiology: clinical correlation with a human patient simulator. *Journal of Clinical Monitoring & Computing*, 16(5-6), pp. 465-470.

26 Ewy, G.A., Felner, J.M., Juul, D., Mayer, J.W., Sajid, A.W., & Waugh, R.A. (1987) Test of a cardiology patient simulator with students in fourth-year electives. *Journal of Medical Education*, 62(9), pp. 738-743.

27 Farnsworth, S.T., Tarmage, D., Egan, T.D., Johnson, S.E., & Westenskow, D. (2000) Teaching sedation and analgesia with simulation. *Journal of Clinical Monitoring & Computing*, 16(4), pp. 273-285.

28 Ferlitsch, A., Glauninger, P., Gupper, A., Schillinger, M., Haefner, M., Gangl, A., & Schoefl R. (2002) Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. *Endoscopy*, 34(9), pp. 698-702.

29 Forrest, F.C., Taylor, M.A., Postlethwaite, K., & Aspinall, R. (2002) Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *British Journal of Anaesthesia*, 88(3), pp. 338-344.

30 Fried, G.M., Derosis, A.M., Bothwell, J., & Sigman, H.H. (1999) Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. *Surgical Endoscopy*, 13, pp. 1077-1081.

31 From, R.P., Pearson, K.S., Albanese, M.A., Moyers, J.R., Sigurdsson, S.S., & Dull, D.L. (1994) Assessment of an interactive learning system with "sensorized" manikin head for airway management instruction. *Anesthesia & Analgesia,* 79(1), pp. 136-142.

32 Gallagher, A.G., & Satava, R.M. (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. *Surgical Endoscopy*, 16(12), pp. 1746-1752.

33 Gallagher, A.G., Hughes, C., McClure, N., & McGuigan, J. (2000) A case-control comparison of traditional and virtual reality training in laparoscopic psychomotor performance. *Minimally Invasive Therapeutics & Allied Technology*, 9, pp. 347-352.

34 Gallagher, A.G., Reinhardt-Rutland, A.H., McGuigan, J., Crothers, I., Browning J. & McClure, N. (1999) Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy*, 31(4), pp. 310-313.

35 Gass, D.A., & Curry, L. (1983) Physicians' and nurses' retention of knowledge and skill after training in cardiopulmonary resuscitation. *Canadian Medical Association Journal*, 128(5), pp. 550-551.

36 Gerson, L.B., & Van Dam, J. (2003) A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy*, 35(7), pp. 569-575.

37 Gilbart, M.K., Hutchison, C.R., Cusimano, M.D., & Regehr, G. (2000) A computer-based trauma simulator for teaching trauma management skills. *The American Journal of Surgery*, 179(3), pp. 223-228.

38 Gor, M., McCloy, R., Stone, R., & Smith, A. (2003) Virtual reality laparoscopic simulator for assessment in gynaecology. *British Journal of Obstetrics & Gynaecology*, 110(2), pp. 181-187.

39 Gordon, M.S., Ewy, G.A., DeLeon, A.C. Jr., Waugh, R.A., Felner, J.M., Forker, A.D., Gessner, I.H., Mayer, J.W., & Patterson, D. (1980) "Harvey," the cardiology patient simulator: pilot studies on teaching effectiveness. *The American Journal of Cardiology*, 45(4), pp. 791-796.

40 Grantcharov, T.P., Bardram, L., Funch-Jensen, P., & Rosenberg, J. (2003) Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *The American Journal of Surgery*, 185(2), pp. 146-149.

41 Hamilton, E.C., Scott, D.J., Fleming, J.B., Rege, R.V., Laycock, R., Bergen, P.C., Tesfday, S.T., & Jones, D.B. (2002) Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy*, 16, pp. 406-411.

42 Hamilton, E.C., Scott, D.J., Kapoor, A., Nwariaku, F., Bergen, P.C., Rege, R.V., Tesfay, S.T., & Jones, D.B. (2001) Improving operative performance using a laparoscopic hernia simulator. *The American Journal of Surgery*, 182(6), pp. 725-728.

43 Hasson, H.M., Kumari, N.V.A., & Eekhout, J. (2001) Training simulator for developing laparoscopic skills. *Journal of the Society of Laparoendoscopic Surgeons*, 5, pp. 255-265.

44 Hosking, E.J. (1998) Does practicing intubation on a manikin improve both understanding and clinical performance of the task by medical students. *Anesthesia Points West*, 31(2), pp. 25-28.

45 Hyltander, A., Liljegren, E., Rhodin, P.H., & Lonroth, H. (2002) The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surgical Endoscopy*, 16(9), pp. 1324-1328.

46 Issenberg, S.B., McGaghie, W.C., Gordon, D.L., Symes, S., Petrusa, E.R., Hart, I.R., & Harden, RM. (2002) Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching & Learning in Medicine*, 14(4), pp. 223-228.

47 Jordan, J.A., Gallagher, A.G., McGuigan, J., & McClure, N. (2001) Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surgical Endoscopy*, 15, pp. 1080-1084.

48 Jordan JA, Gallagher AG, McGuigan J, McGlade K, & McClure N. (2000) A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. *The American Journal of Surgery*, 180(3), pp. 208-211.

49 Kaczorowski, J., Levitt, C., Hammond, M., Outerbridge, E., Grad, R., Rothman, A., & Graves, L. (1998) Retention of neonatal resuscitation skills and knowledge: a randomized controlled trial. *Family Medicine*, 30(10), pp. 705-711.

50 Katz, R., Nadu, A., Olsson, L.E., Hoznek, A., de la Taille, A., Salomon, L., & Abbou, CC. A simplified 5-step model for training laparoscopic urethrovesical anastomosis. *Journal of Urology*, 169(6), pp. 2041-2044.

51 Kothari, S.N., Kaplan, B.J., DeMaria, E.J., Broderick, T.J., & Merrell, R.C. (2002) Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 12(3), pp. 167-173.

52 Kovacs, G., Bullock, G., Ackroyd-Stolarz, S., Cain, E., & Petrie D. A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Annals of Emergency Medicine*, 36(4), pp. 301-309.

53 Mackay, S., Morgan, P., Datta, V., Chang, A., & Darzi, A. (2002) Practice distribution in procedural skills training: A randomized controlled trial. *Surgical Endoscopy*, 16(6), pp. 957-961.

54 MacDonald, J., Williams, R.G., & Rogers, D.A. (2003) Self-assessment in simulation-based surgical skills training. *The American Journal of Surgery*, 185(4), pp. 319-322.

55 MacMillan, A.I.M., & Cuschieri, A. (1999) Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: predictive and concurrent validity. *The American Journal of Surgery*, 177(3), pp. 274-277.

56 Marshall, R.L., Smith, J.S., Gorman, P.J., Krummel, T.M., Haluck, R.S., & Cooney, R.N. (2001) Use of a human patient simulator in the development of resident trauma management skills. *The Journal of Trauma Injury, Infection & Critical Care*, 51(1), pp. 17-21.

57 Modell, J.H., Cantwell, S., Hardcastle, J., Robertson, S., & Pablo, L. (2002) Using the human patient simulator to educate students of veterinary medicine. *Journal of Veterinary Medical Education*, 29(2), pp. 111-116.

58 Monsky, W.L., Levine, D, Mehta, T.S., Kane, R.A., Ziv, A., Kennedy, B., & Nisenbaum, H. (2002) Using a sonographic simulator to assess residents before overnight call. *American Journal of Roentgenology*, 178(1), pp. 35-39.

59 Moorthy, K., Smith, S., Brown, T., Bann, S., & Darzi, A. (2003) Evaluation of virtual reality bronchoscopy as a learning and assessment tool. *Respiration*, 70(2), pp. 195-199.

60 Morgan, P.J., Cleave-Hogg, D., McIlroy, J., & Devitt, J.H. (2002) A comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology*, 96(1), pp. 10-16.

61 Nadel, F.M., Lavelle, J.M., Fein, J.A., Giardino, A.P., Decker, J.M., & Durbin, D.R. (2000) Teaching resuscitation to pediatric residents: the effects of an intervention. *Archives of Pediatric Adolescent Medicine*, 154(10), pp. 1049-54.

62 Neumann, M., Hahn, C., Horbach, T., Schneider, I., Meining, A., Heldwein, W., Rosch, T., & Hohenberger, W. (2003) Score card endoscopy: a multicenter study to evaluate learning curves in 1-week courses using the erlangen endo-trainer. *Endoscopy*, 35(6), pp. 515-520.

63 Neumann, M., Stangl, T., Auenhammer, G., Horbach, T., Hohenberger, W., & Schneider, I. (2003) Laparoscopic cholecystectomy. Training on a bio-simulation model with learning success documented using score-cards. *Der Chirurg*, 74(3), pp. 208-213.

64 Neumann, M., Siebert, T., Rausch, J., Horbach, T., Ell, C., Manegold, C., Hohenberger, W., & Schneider, I. (2003) Scorecard endoscopy: a pilot study to assess basic skills in trainees for upper gastrointestinal endoscopy. *Langenbecks Archives of Surgery*, 387(9-10), pp. 386-391.

65 Noordergraaf, G.J., Van Gelder, J.M., Van Kesteren, R.G., Diets, R.F., & Savelkoul, T.J. (1997) Learning cardiopulmonary resuscitation skills: does the type of mannequin make a difference? *European Journal of Emergency Medicine*, 4(4), pp. 204-209.

66 Nyssen, A.S., Larbuisson, R., Janssens, M., Pendeville, P., & Mayne, A. (2002) A comparison of the training value of two types of anesthesia simulators: computer screen-based and mannequin-based simulators. *Anesthesia & Analgesia*, 94(6), pp. 1560-1565.

67 Oddone, E.Z., Waugh, R.A., Samsa, G., Corey, R., & Feussner J.R. (1993) Teaching cardiovascular examination skills: results from a randomized controlled trial. *The American Journal of Medicine,* 95(4), pp. 389-396.

68 Olympio, M.A., Whelan, R., Ford, R.P., & Saunders, I.C. (2003) Failure of simulation training to change residents' management of oesophageal intubation. *British Journal of Anaesthesia*, 91(3), pp. 312-318.

69 Ost, D., DeRosiers, A., Britt, E.J., Fein, A.M., Lesser, M.L., & Mehta, A.C. (2001) Assessment of a bronchoscopy simulator. *American Journal of Respiratory Critical Care Medicine*, 164(12), pp. 2248-2255.

70 O'Toole, R.V., Playter, R.R., Krummel, T.M., Blank, W.C., Cornelius, N.H., Roberts, W.R., Bell, W.J., & Raibert, M. (1999) Measuring and developing suturing technique with a virtual reality surgical simulator. *Journal of the American College of Surgeons*, 189(1), pp. 114-128.

71 Ovassapian, A., Yelich, S.J., Dykes, M.H., & Golman, M.E. (1988) Learning fiberoptic intubation: use of simulators v. traditional teaching. *British Journal of Anaesthesia*, 61(2), pp. 217-220.

72 Owen, H., & Plummer, J.L. (2002) Improving learning of a clinical skill: the first year's experience of teaching endotracheal intubation in a clinical simulation facility. *Medical Education*, 36(7), pp. 635-642.

73 Pearson, A.M., Gallagher, A.G., Rosser, J.C., & Satava, R.M. (2002) Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. *Surgical Endoscopy*, 16, pp. 130-137.

74 Peugnet, F., Dubois, P., & Rouland, J.F. (1998) Virtual reality versus conventional training in retinal photocoagulation: a first clinical assessment. *Computer Aided Surgery*, 3(1), pp. 20-26.

75 Pittini, R., Oepkes, D., Macrury, K., Reznick, R., Beyene, J., & Windrim, R. (2002) Teaching invasive perinatal procedures: assessment of a high fidelity simulator-based curriculum. *Ultrasound Obstetrics & Gynecology*, 19(5), pp. 478-483.

76 Powers, T.W., Murayama, K.M., Toyama, M., Murphy, S., Denham, E.W. 3rd, Derossis, A.M., & Joehl, R.J. (2002) Housestaff performance is improved by participation in a laparoscopic skills curriculum. *The American Journal of Surgery*, 184(6), pp. 626-629, discussion 629-630.

77 Prystowsky, J.B., Regehr, G., Rogers, D.A., Loan, J.P., Hiemenz, L.L., & Smith, K.M. (1999) A virtual reality module for intravenous catheter placement. The American Journal of Surgery, 177(2), pp. 171-175.

78 Pugh, C.M., Srivastava, S., Shavelson, R., Walker, D., Cotner, T., Scarloss, B., Kuo, M., Rawn, C., Dev, P., Krummel, T.H., & Heinrichs, L.H. (2001) The effect of simulator use on learning and self-assessment: the case of Stanford University's E-Pelvis simulator. *Studies in Health Technology & Informatics*, 81, pp. 396-400.

79 Risucci, D., Cohen, J.A., Garbus, J.E., Goldstein, M., & Cohen, M.G. (2001) The effects of practice and instruction on speed and accuracy during resident acquisition of simulated laparoscopic skills. *Current Surgery*, 58(2), pp. 230-235.

80 Risucci, D., Geiss, A., Gellman, L., Pinard, B., & Rosser J.C. (2000) Experience and visual perception in resident acquisition of laparoscopic skills. *Current Surgery*, 57(4), pp. 368-372.

81 Rowe, R., & Cohen, R.A. (2002) An evaluation of a virtual reality airway simulator. *Anesthesia & Analgesia*, 95(1), pp. 62-66.

82 Salen, P., O'Connor, R., Passarello, B., Pancu, D., Melanson, S., Arcona, S., & Heller, M. (2001) Fast education: a comparison of teaching models for trauma sonography. *The Journal of Emergency Medicine*, 20(4), pp. 421-425.

83 Salvendy, G., Root, C.M., Schiff, A.J., Cunningham, P.R., & Ferguson, G.W. (1975) A second generation training simulator for acquisition of psychomotor skills in cavity preparation. *Journal of Dental Education*, 39(7), pp. 466-471.

84 Scherer, Y.K., Bruce, S.A., Graves, B.T., & Erdley, W.S. (2003) Acute care nurse practitioner education: enhancing performance through the use of clinical simulation. *American Association of Critical Care Nurses Clinical Issues*, 14(3), pp. 331-341.

85 Schwid, H.A., Rooke, G.A., Carline, J., Steadman, R.H., Murray, W.B., Olympio, M., Tarver, S., Steckner, K., & Wetstone, S. (2002) Anesthesia Simulator Research Consortium. Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study. *Anesthesiology*, 97(6), pp. 1434-1444.

86 Scott, D.J., Young, W.N., Tesfay, S.T., Frawley, W.H., Rege, R.V., & Jones, D.B. (2001) Laparosopic skills training. *The American Journal of Surgery*, 182, pp. 137-142.

87 Scott, D.J., Rege, R.V., Bergen, P.C., Guo,W.A., Laycock, R., Tesfay, S.T., Valentine, R.J., & Jones, D.B. (2000) Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 10(4), pp. 183-190.

88 Scott, D.J., Bergen, P.C., Rege, R.V., Laycock, R., Tesfay, S.T., Valentine, R.J., Euthus, D.M., Jeyarajah, D.R., Thompson, W.M., & Jones, D.B. (2000) Laparoscopic training on bench models: better and more cost effective than operating room experience? *Journal of the American College of Surgeons*, 191, pp. 272-283.

89 Sedlack, R., Petersen, B., Binmoeller, K., & Kolars, J. (2003) A direct comparison of ERCP teaching models. *Gastrointestinal Endoscopy*, 57(7), pp. 886-890.

90 Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., & Satava, R.M. (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of Surgery*, 236(4), pp. 458-463.

91 Shah, J., Montgomery, B., Langley, S., & Darzi, A. (2002) Validation of a flexible cystoscopy course. *British Journal of Urology International*, 90(9), pp. 833-835.

92 Shah, J., & Darzi, A. (2002) Virtual reality flexible cystoscopy: a validation study. *British Journal of Urology International,* 90(9), pp. 828-832.

93 Stratton, S.J., Kane, G., Gunter, C.S., Wheeler, N.C., Ableson-Ward, C., Reich, E., Pratt, F,D., Ogata, G., & Gallagher, C. (1991) Prospective study of manikin-only versus manikin and human subject endotracheal intubation training of paramedics. *Annals of Emergency Medicine*, 20(12), pp. 1314-1318.

94 Strom, P., Kjellin, A., Hedman, L., Johnson, E., Wredmark, T., & Fellander-Tsai, L. (2003) Validation and learning in the Procedicus KSA virtual reality surgical simulator. *Surgical Endoscopy*, 17(2), pp. 227-231.

95 Tan, G.M., Ti, L.K., Suresh, S., Ho, B.S., & Lee, T.L. (2002) Teaching first-year medical students physiology: does the human patient simulator allow for more effective teaching? *Singapore Medical Journal*, 43(5), pp. 238-242.

96 Treloar, D., Hawayek, J., Montgomery, J.R., & Russell, W: Medical Readiness Trainer Team. (2001) On-site and distance education of emergency medicine personnel with a human patient simulator. *Military Medicine*, 166(11), pp. 1003-1006.

97 Torkington, J., Smith, S.G.T., Rees, B.I., & Darzi, A. (2001) The role of the basic surgical course in the acquisition and retention of laparoscopic skill. *Surgical Endoscopy*, 15, pp. 1071-1075.

98 Torkington, J., Smith, S.G.T., Rees, B.I., Darzi, A. (2001) Skill transfer from virtual reality to a real laparoscopic task. *Surgical Endoscopy*, 15, pp. 1076-1079.

99 Tuggy, M.L. (1998) Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *Journal of the American Board of Family Practice*, 11(6), pp. 426-433.

100 Watterson, J.D., Beiko, D.T., Kuan, J.K., & Denstedt, J.D. (2002) Randomized prospective blinded study validating acquisition of ureteroscopy skills using computer based virtual reality endourological simulator. *Journal of Urology*, 168(5), pp. 1928-1932.

101 Weghorst, S., Airola, C., Oppenheimer, P., Edmond, C.V., Patience, T., Heskamp, D., & Miller, J. (1998) Validation of the Madigan ESS simulator. *Studies in Health Technology & Informatics*, 50, pp. 399-405.

102 Wik, L., Dorph, E., Auestad, B., & Steen, P.A. (2003) Evaluation of a defibrillator-basic cardiopulmonary resuscitation programme for non medical personnel. *Resuscitation*, 56(2), pp. 167-172.

103 Wik, L., Myklebust, H., Auestad, B.H., & Steen, P.A. (2002) Retention of basic life support skills 6 months after training with an automated voice advisory manikin system without instructor involvement. *Resuscitation*, 52(3), pp. 273-279.

104 Wik, L., Thowsen, J., & Steen, P.A. (2001) An automated voice advisory manikin system for training in basic life support without an instructor. A novel approach to CPR training. *Resuscitation*, 50(2), pp. 167-172.

105 Wilhelm, D.M., Ogan, K., Roehrborn, C.G., Cadeddu, J.A., & Pearle, M.S. (2002) Assessment of basic endoscopic performance using a virtual reality simulator. *Journal of the American College of Surgery*, 195(5), pp. 675-681.

106 Wong, D.T., Prabhu, A.J., Coloma, M., Imasogie, N., & Chung, F.F. (2003) What is the minimum training required for successful cricothyroidotomy?: a study in mannequins. *Anesthesiology*, 98(2), pp. 349-353.

107 Woolliscroft, J.O., Calhoun, J.G., Tenhaken, J.D., & Judge, R.D. (1987) Harvey: the impact of a cardiovascular teaching simulator on student skill acquisition. *Medical Teacher*, 9(1), pp. 53-57.

108 Yoshii, C., Anzai, T., Yatera, K., Kawajiri, T., Nakashima, Y., & Kido, M. (2002) A new medical education using a lung sound auscultation simulator called “Mr. Lung”. *Journal UOEH*, 24(3), pp. 249-255.

109 Young, T.J., Hayek, R., & Philipson, S.A. (1998) A cervical manikin procedure for chiropractic skills development. *Journal of Manipulative & Physiological Therapeutics*, 21(4), pp. 241-245

Appendix 3 – Comparative studies with simulator used only as assessment intervention

1 Bucx, M.J., van Geel, R.T., Wegener, J.T., Robers, C., & Stijnen, T. (1995) Does experience influence the forces exerted on maxillary incisors during laryngoscopy? A manikin study using the Macintosh laryngoscope. *Canadian Journal of Anaesthesia*, 42(2), pp. 144-149.

2 Byrne, A.J., & Jones, J.G. (1997) Responses to simulated anaesthetic emergencies by anaesthetists with different durations of clinical experience. *British Journal of Anaesthesia*, 78(5), pp. 553-536.

3 Datta, V., Mandalia, M., Mackay, S., & Darzi, A. (2002) The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. *Surgical Endoscopy*, 16(10), pp. 1459-1463.

4 DeAnda, A., & Gaba, D.M. (1991) Role of experience in the response to simulated critical incidents. *Anesthesia & Analgesia*, 72(3), pp. 308-315.

5 Derosis, A.M., Antoniuk, M., & Fried, G.M. (1999) Evaluation of laparoscopic skills: a 2-year follow-up during residency training. *Canadian Journal of Surgery*, 42(4), pp. 293-296.

6 Devitt, J.H., Kurrek, M.M., Cohen, M.M., & Cleave-Hogg, D. (2001) The validity of performance assessments using simulation. *Anesthesiology*, 95(1), pp. 36-42.

7 Dorafshar, A.H., O'Boyle, D.J., & McCloy, R.F. (2002) Effects of a moderate dose of alcohol on simulated laparoscopic surgical performance. *Surgical Endoscopy*, 16(12), pp. 1753-1758.

8 Eastridge, B.J., Hamilton, E.C., O'Keefe, G.E., Rege, R.V., Valentine, R.J., Jones, D.J., Tesfay, S., & Thal, E.R. (2003) Effect of sleep deprivation on the performance of simulated laparoscopic surgical kill. *The American Journal of Surgery*, 186(2), pp. 169-174.

9 Emam, T.A., Hanna, G.B., Kimber, C., & Cuschieri, A. (2000) Differences between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. *Digestive Surgery*, 17, pp. 120-125.

10 Francis, N.K., Hanna, G.B., & Cuschieri A. (2002) The performance of master surgeons on the advanced Dundee endoscopic psychomotor tester: contrast validity study. *Archives of Surgery*, 137(7), pp. 841-844.

11 Fraser, S.A., Klassen, D.R., Feldman, L.S., Ghitulescu, G.A., Stanbridge, D., & Fried, G.M. (2003) Evaluating laparoscopic skills. *Surgical Endoscopy*, 17(6), pp. 964-947.

12 Gaba, D.M., & DeAnda, A. (1989) The response of anesthesia trainees to simulated critical incidents. *Anesthesia & Analgesia*, 68(4), pp. 444-451.

13 Gallagher, A.G., Richie, K., McClure, N., & McGuigan, J. (2001) Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World Journal of Surgery*, 25(11), pp. 1478-1483.

14 Gallagher, H.J., Allan, J.D., & Tolley, D.A. (2001) Spatial awareness in urologists:are they different? *British Journal of Urology International*, 88, pp. 666-670.

15 Grantcharov, T.P., Bardram, L., Funch-Jensen, P., & Rosenberg, J. (2003) Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. *Surgical Endoscopy*, 17(7), pp. 1082-1085.

16 Grantcharov, T.P., Bardman, L., Funch-Jensen, P., & Rosenberg, J. (2001) Laparoscopic performance after one night on call in a surgical department: prospective study. *British Medical Journal*, 323, pp. 1222-1223.

17 Haluck, R.S., Webster, R.W., Snyder, A.J., Melkonian, M.G., Mohler, B.J., Dise, M.L., & Lefever, A. (2001) A virtual reality surgical trainer for navigation in laparoscopic surgery. *Studies in Health Technology& Informatics*, 81, pp. 171-177.

18 Hanna, G.B., Cresswell, A.B., & Cuschieri, A. (2002) Shadow depth cues and endoscopic task performance. *Archives of Surgery*, 137, pp. 1166-1169.

19 Holcomb, J.B., Dumire, R.D., Crommett, J.W., Stamateris, C.E., Fagert, M.A., Cleveland, J.A., Dorlac, G.R., Dorlac, W.C., Bonar, J.P., Hira, K., Aoki, N., & Mattox, K.L. (2002) Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *The Journal of Trauma Injury, Infection & Critical Care*, 52(6), pp. 1078-1085.

20 Hotchkiss, M.A., Biddle, C., & Fallacaro, M. (2002) Assessing the authenticity of the human simulation experience in anesthesiology. *American Association of Nurse Anesthetists Journal*, 70(6), pp. 470-473.

21 Howard, S.K., Gaba, D.M., Smith, B.E., Weinger, M.B., Herndon, C., Keshavacharya, S., & Rosekind, M.R. (2003) Simulation study of rested versus sleep-deprived anesthesiologists. *Anesthesiology,* 98(6), pp. 1345-1355.

22 Johnston, R., Bhoyrul, S., Way, L., Satava, R., McGovern, K., Fletcher, J.D., Rangel, S., & Loftin, R.B. (1996) Assessing a virtual reality surgical skills simulator. *Studies in Health Technology & Informatics*, 29, pp. 608-617.

23 Jones, D.B., Brewer, J.D., & Soper, N.J. (1996) The influence of three-dimensional video systems on laparoscopic task performance. *Surgical Laparoscopy & Endoscopy*, 6(3), pp. 191-197.

24 Jones, J.S., Hunt, S.J., Carlson, S.A., & Seamon, J.P. (1997) Assessing bedside cardiologic examination skills using "Harvey," a cardiology patient simulator. *Academic Emergency Medicine*, 4(10), pp. 980-985.

25 Lampotang, S. (1998) Influence of pulse oximetry and capnography on time to diagnosis of critical incidents in anesthesia: a pilot study using a full-scale patient simulator. *Journal of Clinical Monitoring & Computing*, 14(5), pp. 313-321.

26 Mackay, S., Datta, V., Chang, A., Shah, J., Kneebone, R., & Darzi, A. (2003) Multiple Objective Measures of Skill (MOMS): A new approach to the assessment of technical ability in surgical trainees. *Annals of Surgery*, 238(2), pp. 291-300.

27 MacDonald, J., Ketchum, J., Williams, R.G., & Rogers, L.Q. (2003) A lay person versus a trained endoscopist. Can the PreOp Endoscopy simulator detect a difference? *Surgical Endoscopy*, 17(6), pp. 896-898.

28 Mahmood, T., & Darzi, A. (2003) A study to validate the colonoscopy simulator. *Surgical Endoscopy*, 17(10), pp. 1583-1589.

29 McCarthy, A., Harley, P., & Smallwood, R. (1999) Virtual arthroscopy training: do the "virtual skills" developed match the real skills required? *Studies in Health Technology & Informatics*, 62, pp. 221-227.

30 McNatt, S.S., & Smith, C.D. (2001) A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. *Surgical Endoscopy*, 15, pp. 1085-1089.

31 Moorthy, K., Munz, Y., Dosis, A., Bann, S., & Darzi, A. (2003) The effect of stress-inducing conditions on the performance of a laparoscopic task. *Surgical Endoscopy*, 17(9), pp. 1481-1484.

32 Morgan, P.J., & Cleave-Hogg, D. (2002) Comparison between medical students' experience, confidence and competence. *Medical Education*, 36(6), pp. 534-539.

33 Morgan, P.J., & Cleave-Hogg, D. (2000) Evaluation of medical students' performance using the anaesthesia simulator. *Medical Education*, 34(1), pp. 42-45.

34 Murray, D., Boulet, J., Ziv, A., Woodhouse, J., Kras, J., & McAllister, J. (2002) An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. *Medical Education*, 36(9), pp. 833-841.

35 Nakajima, K., Wasa, M., Takiguchi, S., Taniguchi, E., Soh, H., Ohashi, S., & Okada, A. (2003) A modular laparoscopic training program for pediatric surgeons. *Journal of the Society of Laparoendoscopic Surgeons*, 7(1), pp. 33-37.

36 Neumann, M., Friedl, S., Meining, A., Egger, K., Heldwein, W., Rey, J.F., Hochberger, J., Classen, M., Hohenberger, W., & Rosch, T. (2002) A score card for upper GI endoscopy: evaluation of interobserver variability in examiners with various levels of experience. *Z Gastroenterology*, 40(10), pp. 857-862.

37 Paisley, A.M., Baldwin, P.J., & Paterson-Brown, S. (2001) Validity of surgical simulation for the assessment of operative skill. *British Journal of Surgery*, 88, pp. 1525-1532.

38 Pedowitz, R.A., Esch, J., & Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy*, 18(6), pp. 1-6.

39 Pichichero, M.E., & Poole, M.D. (2001) Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Archives of Pediatric Adolescent Medicine*, 155(10), pp. 1137-1142.

40 Pugh, C.M., & Youngblood, P. (2002) Development and validation of assessment measures for a newly developed physical examination simulator. *Journal of the American Medical Informatics Association*, 9(5), pp. 448-460.

41 Reznek, M.A., Rawn, C.L., & Krummel, T.M. (2002) Evaluation of the educational effectiveness of a virtual reality intravenous insertion simulator. *Academic Emergency Medicine*, 9(11), pp. 1319-1325.

42 Reznick, R., Regehr, G., Macrae, H., Martin, J., & McCulloch, W. (1996) Testing technical skill via an innovative bench station examination. *The American Journal of Surgery*, 172, pp. 226-230.

43 Rissucci, D., Geiss, A., Gellman, L., Pinard, B., & Rosser, J. (2001) Surgeon-specific factors in the acquisition of laparoscopic surgical skills. *The American Journal of Surgery*, 181(4), pp. 289-293.

44 Rogers, P.L., Jacob, H., Rashwan, A.S., & Pinsky, M.R. (2001) Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. *Critical Care Medicine*, 29(6), pp. 1268-1273.

45 Schijven, M., & Jakimowicz, J. (2003) Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. *Surgical Endoscopy*, 17(5), 803-810.

46 Schijven, M., & Jakimowicz, J. (2002) Face-, expert and referent validity of the Xitact LS500 laparoscopy simulator. *Surgical Endoscopy*, 16(12), pp. 1764-1770.

47 Schijven, M.P., Jakimowicz, J., & Schot, C. (2002) The advanced Dundee endoscopic psychomotor tester (ADEPT) objectifying subjective psychomotor test performance. *Surgical Endoscopy*, 16(6), pp. 943-948.

48 Schwid, H.A., Rooke, G.A., Michalowski, P., & Ross, B.K. (2001) Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. *Teaching & Learning in Medicine*, 13(2), pp. 92-96.

49 Sedlack, R.E., & Kolars, J.C. (2003) Validation of a computer-based colonoscopy simulator. *Gastrointestinal Endoscopy*, 57(2), pp. 214-218.

50 SEMPLE, M., & COOK, R. (2001) Social influence and the recording of blood pressure by student nurse: an experimental study. *Nurse Researcher*, 8(3), pp. 60-71.

51 SHAH, J., BUCKLEY, D., FRISBY, J., & DARZI, A. (2003) Depth cue reliance in surgeons and medical students. *Surgical Endoscopy*, 17(9), pp. 1472-1474.

52 SHAH, J., PAUL, I., BUCKLEY, D., DAVIS, H., FRISBY, J.P., & DARZI, A. (2003) Can tonic accommodation predict surgical performance? *Surgical Endoscopy*, 17(5), pp. 787-790.

53 SHERMAN, K.P., WARD, J.W., WILLS, D.P., SHERMAN, V.J., & MOHSEN, A.M. (2001) Surgical trainee assessment using a VE knee arthroscopy training system (VE-KATS): experimental results. *Studies in Health Technology & Informatics*, 81, pp. 465-470.

54 SORRENTO, A., & PICHICHERO, M.E. (2001) Assessing diagnostic accuracy and tympanocentesis skills by nurse practitioners in management of otitis media. *Journal of the American Academy of Nurse Practitioner*s, 13(11), pp. 524-529.

55 ST CLAIR, E.W., ODDONE, E.Z., WAUGH, R.A., COREY, G.R., & FEUSSNER, J.R. (1992) Assessing housestaff diagnostic skills using a cardiology patient simulator. *Annals of Internal Medicine*, 117(9), pp. 751-756.

56 SUNG, W.H., FUNG, C.P., CHEN, A.C., YUAN, C.C., NG, H.T., & DOONG, J.L. (2003) The assessment of stability and reliability of a virtual reality-based laparoscopic gynecology simulation system. *European Journal of Gynaecological Oncology*, 24(2), pp. 143-146.

57 TAFFINDER, N., SUTTON, C., FISHWICK, R.J., MCMANUS, I.C., & DARZI A. (1998) Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomized controlled studies using the MIST VR laparoscopic simulator. *Studies in Health Technology & Informatics*, 50, pp. 124-130.

58 TWIGG, S.J., MCCORMICK, B., & COOK, T.M. (2003) Randomized evaluation of the performance of single-use laryngoscopes in simulated easy and difficult intubation. *British Journal of Anaesthesia*, 2003, 90(1), pp. 8-13.

59 UCHAL, M., BROGGER, J., RUKAS, R., KARLSEN, B., & BERGAMASCHI, R. (2002) In-line versus pistol-grip handles in a laparoscopic simulator. A randomized controlled crossover trial. *Surgical Endoscopy*, 16(12), pp. 1771-1773.

60 UHRICH, M.L., UNDERWOOD, R.A., STANDEVEN, J.W., SOPER, N.J., & ENGSBERG, J.R. (2002) Assessment of fatigue, monitor placement, and surgical experience during simulated laparoscopic surgery. *Surgical Endoscopy*, 16(4), pp. 635-639.

61 WENTINK, M., BREEDVELD, P., STASSEN, L.P., OEI, I.H., & WIERINGA, P.A. (2002) A clearly visible endoscopic instrument shaft on the monitor facilitates hand-eye coordination. *Surgical Endoscopy*, 16(11), pp. 1533-1537.

62 WESTMAN, E.C., MATCHAR, D.B., SAMSA, G.P., MULROW, C.D., WAUGH, R.A., & FEUSSNER, J.R. (1995) Accuracy and reliability of apical S3 gallop detection. *Journal of General Internal Medicine*, 10(8), pp. 455-457.

Notes

Notes