Effect of practice on standardised learning outcomes in simulation-based medical education

William C McGaghie,1 S Barry Issenberg,2 Emil R Petrusa3 & Ross J Scalese2

OBJECTIVES This report synthesises a subset of 31 journal articles on high-fidelity simulation-based medical education containing 32 research studies drawn from a larger qualitative review published previously. These studies were selected because they present adequate data to allow for quantitative synthesis. We hypothesised an association between hours of practice in simulation-based medical education and standardised learning outcomes measured as weighted effect sizes.

METHODS Journal articles were screened using 5 exclusion and inclusion criteria. Response data were extracted and 3 judges independently coded each study. Learning outcomes were standardised using a common metric, the average weighted effect size (AWES), due to the heterogeneity of response measures in individual studies. ANOVA was used to evaluate AWES differences due to hours of practice on a high-fidelity medical simulator cast in 5 categories. The eta squared ($\eta^2$) statistic was used to assess the association between AWES and simulator practice hours.

RESULTS There is a strong association ($\eta^2 = 0.46$) between hours of practice on high-fidelity medical simulators and standardised learning outcomes. The association approximates a dose–response relationship.

CONCLUSIONS Hours of high-fidelity simulator practice have a positive, functional relationship with standardised learning outcomes in medical education. More rigorous research methods and more stringent journal editorial policies are needed to advance this field of medical education research.

KEYWORDS education, medical, undergraduate/*methods; teaching/*methods; United States; learning; feedback, psychological; teaching materials; clinical competence/*standards.

Medical Education 2006; 40: 792–797
doi:10.1111/j.1365-2929.2006.02528.x

INTRODUCTION

A recently published qualitative, systematic review spanning 34 years and covering 670 peer-reviewed journal articles identified 10 features and uses of high-fidelity medical simulations in a range of specialties, including anaesthesiology, cardiology and surgery, that lead to effective learning.1 The features and uses, in order of importance, are listed here.

1 Feedback is provided during the learning experience. Knowledge of results about one’s performance stands out from the research legacy as the most important feature of simulation-based medical education in promoting effective learning.

2 Learners engage in repetitive practice. This involved focused and repetitive learner engagement in practice sessions where the intent is skill improvement, not just idle repetition.

3 The simulator is integrated into the medical curriculum. Simulation-based educational experiences are a routine feature of the normal educational schedule and are grounded in learner performance evaluation.

4 Learners practise with increasing levels of difficulty. Learners engage in medical skills practice across a range of difficulty levels, beginning with basics and advancing to progressively higher difficulty levels based on objective measurements.
Overview

What is already known on this subject

A qualitative research synthesis published previously suggests that at least 10 features and uses of high-fidelity medical simulations lead to effective learning.

What this study adds

Quantitative research synthesis was used on a subset of the studies previously reviewed qualitatively. This study tests the hypothesis that there is a quantitative relationship between hours of simulator practice and the magnitude of learning gains. Results show that hours of practice on high-fidelity medical simulations have a strong empirical association with standardised learning outcomes.

Suggestions for further research

Future studies should address the putative dose–response relationship between simulator practice hours and learning outcomes. Increased rigour is needed to advance research in this field.

5 The simulator is adaptable to multiple learning strategies. Simulation-based learning strategies included but were not limited to instructor-centred formats, small-group tutorials and independent study, and completely independent learning, depending on the learning objectives being addressed.

6 The simulator captures clinical variation. Simulators that can represent a wide variety of patient problems provide more sampling than simulators that cover a narrow patient range.

7 The simulator is embedded in a controlled environment. High-fidelity simulations work best in controlled educational settings where learners can make, detect and correct patient care errors without negative consequences, unlike in real clinical environments.

8 The simulator permits individualised learning. Educational experiences can be adapted in a reproducible and standardised way to individualised learner needs so that learners are active participants, not passive bystanders.

9 Learning outcomes are clearly defined and measured. Educational goals have tangible, objective measures that document learner progress in terms of training benchmarks.

10 The simulator is a valid (high-fidelity) approximation of clinical practice. The simulation and the behaviour it provokes come close to, but never exactly duplicate, clinical challenges that happen in genuine patient care contexts.

The original goal of the review was to perform a quantitative meta-analysis of this literature. However, the heterogeneity of research designs, educational interventions, outcome measures and journal-specific statistical reporting conventions precluded data synthesis using meta-analysis. Consequently, the investigators resorted to a qualitative, systematic review of the literature on high-fidelity simulations in medical education as the best available approach given the condition of the current published research, as advocated elsewhere by Reed and colleagues.

We determined in retrospect that the direct effects on learning from 1 of the 10 simulation use variables, learners engage in repetitive practice, was the only variable that could be evaluated quantitatively from a subset of 31 research reports from the larger journal article pool. Repetitive practice on a simulator was second in frequency among the 10 variables. Such a quantitative research synthesis can only be conducted when the different learning outcome measures contained in individual research studies can be cast on a common, standardised metric. Repetitive practice means that learners work with the medical simulator over multiple trials toward the goal of skill acquisition or improvement. In this report, we test the hypothesis that there is a quantitative association between the hours of repetitive practice in simulation-based medical education and standardised cognitive and psychomotor learning outcomes in a variety of disciplines.

METHODS

Data retrieval

The previous review used 91 search terms and 5 literature databases to identify the 670 journal articles eligible for review. The 670 articles were screened using 4 exclusion and inclusion criteria, leaving 109 articles for qualitative synthesis. The criteria excluded review articles without original data and included reports where a medical education simulator was used with quantitative outcome measurement.
experimental or quasi-experimental comparative research, and comparative research where the simulator was used as an educational intervention. A fifth inclusion criterion was added for this study: presentation of research methods and results in sufficient detail to enable quantitative synthesis. This is essential because effect size calculation requires at minimum that original data are presented as group means and standard deviations, not just as calculated P-values. Only 31 journal articles reporting 32 research investigations (32/109 = 29%; 32/670 = 5%) survived all 5 exclusion and inclusion criteria for this study. A list of these articles and a tabular presentation of their features (e.g. study design, learner categories, outcome measures) is available online at http://www.blackwell-synergy.com.

Data synthesis

Means, standard deviations, and the number of learners for each outcome variable in the 32 research studies in the final pool were extracted. Each of the studies was coded (blind) by 3 independent judges. Data were extracted from each study to measure the intensity of the educational intervention coded as an ordinal variable,³ hours of simulator practice in 5 categories: 0 = none reported; 1 = 0–1.0 hour (e.g. a 1-hour workshop); 2 = 1.1–3.0 hours (e.g. a half-day of training); 3 = 3.1–8.0 hours (e.g. a full day of training); 4 = 8 + hours (e.g. at least 1 day of training).

Data analysis

Data analysis was carried out in 3 steps. Firstly, the interrater reliability of the 3-judge coding was assessed. Secondly, a weighted effect size (WES) was calculated to standardise each learning outcome variable in the 32 research studies. The WES accounts for the number of cases represented by each variable, an approximation to study power.⁴,⁵ Effect sizes were averaged for studies with multiple outcome variables (AWES) and their heterogeneity (reliability) represented by Iota squared (I² = non-chance variation) was assessed.⁶ This casts the heterogeneous learning outcome variables contained in the 32 research studies onto a common, standardised metric. Random effects models were used for all of these analyses. Thirdly, one-way analysis of variance (ANOVA) was used with research studies as the unit of analysis. The ANOVA determined if hours of simulator practice in the 5 categories (independent variable) was responsible for variation in the AWES (dependent variable) among the 32 studies. Duncan’s multiple range test was used to identify homogeneous research study subsets. The η² statistic evaluated the proportion of variance in AWES explained by the 5 categories of simulator practice hours.

RESULTS

Coding of the 32 research studies for hours of simulator practice achieved complete agreement. This obviates calculation of kappa coefficients.

A graphic portrait and descriptive statistics (mean, standard deviation) for the AWES from the 32 research reports in this study are given in Fig. 1. The AWES values are graphed according to the 5 practice time categories. The number of research studies for each category ranges from 2 to 11 (median = 5). The heterogeneity of the AWES variable (I² = 99.7%) shows that variability across the studies is not due to chance.

One-way ANOVA on AWES using hours of simulator practice as the independent variable yields a highly significant result (F[4,27] = 5.77, P < 0.002). Post hoc analysis using Duncan’s multiple range test indicates 2 homogeneous subsets within the set of 5 simulator practice categories. The first homogeneous subset comprises the first 4 levels of simulator practice hours: none reported to 3.1–8.0. The second homogeneous subset comprises only the 2 studies that involved more than 8.1 hours of medical learner practice on the simulator. Overall, the data show that hours of simulator practice account for almost 50% of the variation in the AWES variable (η² = 0.46). This is a substantial proportion of explained outcome variation from a single independent variable.⁷,⁸

![Figure 1](http://www.blackwell-synergy.com)  
**Figure 1** Average weighted effect size and hours of simulator practice.
DISCUSSION

Two principal findings emerge from this study. First, the evidence is clear from these research studies that repetitive practice involving medical simulations is associated with improved learner outcomes. Simulation-based practice in medical education appears to approximate a dose–response relationship in terms of achieving desired learning outcomes: more practice yields better results. The evidence from the current study pertains to all levels of learners (i.e. students, residents, fellows, senior attending doctors) for a wide variety of clinical specialties (i.e. anaesthesiology, chiropractic, dentistry, emergency medical technicians (EMT), family medicine, internal medicine, nursing, pediatrics, surgery), and for other professionals (i.e. flight attendants). Standardising the learning outcome variable is a strength of the study across the specialties and the different outcome measures in individual research reports.

More recent research studies that were not covered in this quantitative synthesis because they were not included in the earlier systematic review amplify the practice–outcome connection in simulation-based medical education. Two investigations—a randomised trial and a single-group mastery learning study—address the acquisition of advanced cardiac life support (ACLS) skills by internal medicine residents. In this research, skill acquisition is increased under more stringent deliberate practice conditions which involve:

1. intense, repetitive performance of intended cognitive or psychomotor skills in a focused domain; coupled with
2. rigorous skills assessment; that provides learners with
3. specific, informative feedback; that yields increasingly
4. better skills performance in a controlled setting.

The randomised trial demonstrated a 38% increase in residents’ ACLS skills due to a simulation-based, deliberate practice, educational intervention compared with wait-list controls. This finding was replicated with crossover of the randomised groups. The mastery learning study also showed that deliberate practice on a high-fidelity medical simulator produced large and consistent increases in residents’ ACLS skills, although the residents varied in the amount of practice time needed to reach mastery performance standards. Academic aptitude, gauged by USMLE (United States Medical Licensing Examination) Step 1 and 2 scores, had no association with ACLS skill acquisition or mastery outcome measures.

Secondly, few published journal articles on the effectiveness of high-fidelity simulations in medical education have been performed with enough quality and rigour to yield useful results. Only 5% of the research publications in this field meet or exceed the minimum quality standards used for this study. This situation retards scholarly advancement and slows the ability of medical educators to use simulation technology effectively.

We observe from this report and the earlier systematic review that the research literature on high-fidelity simulations in medical education suffers from at least 6 consistent flaws, as listed here.

1. Few published studies display awareness of research and scholarship beyond the scope of individual medical or clinical specialties. Broader knowledge of current research, expressed operationally as inter-specialty journal citations, would improve the state of affairs.
2. Most of the research reports represented in this study have few subjects and give no attention to statistical power.
3. The published work demonstrates a lack of awareness about basic designs used for research in education, and the clinical disciplines.
4. Most current research pays no attention to measurement properties of educational and clinical research variables. This is especially the case regarding the reliability of dependent variables, which was rarely reported in the journal articles reviewed in this study.
5. Statistical reporting conventions are inconsistent across journals in the medical specialties, medical education and the behavioural sciences. At a minimum, journal articles reporting primary research data should present indexes of central tendency (mean, median, mode), dispersion (standard deviation, standard error) and effect size for dependent variables as articulated elsewhere.
6. Properties of educational interventions, such as their strength and integrity, are rarely, if ever, described in reports on simulation-based medical education.

This critique is not only addressed to scholars active in the field of simulation-based investigations in medical education, but also to journal editors who evaluate and publish research. Increased scientific
rigour is clearly needed, as are improved data-reporting conventions. For example, more journal editors need to insist that investigators report descriptive statistics (e.g. means, standard deviations) in addition to the results of statistical tests (e.g. P-values), so that research outcomes can be replicated, synthesised quantitatively, and contribute to cumulative educational science.

This study has several clear limitations. In particular, it is based on a small set of only 31 journal articles that ‘passed scientific muster’ by meeting our 5 screening criteria, and on an uneven number of studies in the 5 simulator practice categories shown in Fig. 1. Although a positive relationship between amount of practice time and size of learning gain seems obvious, the results of this study are strongly influenced by the 2 research reports where practice time met or exceeded 8 hours. The stability of this finding must be verified by further research. However, the magnitude of the F statistic, the low probability that chance alone is responsible for the statistical outcome, the large proportion of outcome variance explained by the statistical model ($\eta^2 = 0.46$), and the robustness of ANOVA from departures of basic statistical assumptions lead us, in sum, to conclude that the results are trustworthy. Further, selected studies only involved high-fidelity simulators. Additional research is needed to determine if similar conclusions are warranted for lower-fidelity simulators. Finally, we recommend strongly that researchers address the 6 methodological weaknesses in future studies.

Contributors: all authors participated in the conception and design of the study. WCM contributed to data collection and analysis. SBI and RJS contributed to data collection. ERP contributed to data analysis. All authors participated in the preparation of the manuscript.

Acknowledgements: we wish to acknowledge Michael S Gordon MD, PhD, Director of the Centre for Research in Medical Education at the University of Miami Miller School of Medicine for his continuous support for our research and scholarship.

Funding: this project was supported by the State of Florida Department of Education and the Friends for Life Volunteer Organization, Miami, Florida, USA.

Conflicts of interest: none.

Ethical approval: not applicable.

REFERENCES


SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article online:

Table S1. Features of included journal articles

Appendix S1. Journal articles included in Average Weighted Effect Size (AWES) calculations

This material is available as part of the online article from http://www.blackwell-synergy.com

Received 7 June 2005; editorial comments to authors 12 September 2005; accepted for publication 20 December 2005