Low- versus high-fidelity simulations in teaching and assessing clinical skills

Fadi Munshi, MD a,*, Hani Lababidi, MD b and Sawsan Alyousef, MD c

a Medical Education Department, College of Medicine — King Fahad Medical City, King Saud bin Abdulaziz University, Riyadh, Kingdom of Saudi Arabia
b CRESENT, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia
c Pediatric Intensive Care Department, Children Hospital, CRESENT, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia

Received 18 December 2014; revised 19 January 2015; accepted 19 January 2015; Available online 23 February 2015

Abstract

Simulation has been widely used in the education of healthcare workers. In simulation training, there is an approximation to reality in which trainees are supposed to react to problems or conditions as they would under genuine circumstances. The educational value of simulations has been determined to be valuable. Simulation has a significant impact on health care education across the disciplines and in both undergraduate and postgraduate studies. Recent development in technologies permits the reproduction of real-life scenarios with acceptable fidelity, thus profoundly enhancing the learning environment. However, the educational outcomes of high- versus low-fidelity simulations remain controversial. This article aims to review the effectiveness of low- and high-fidelity simulations in teaching and assessing clinical skills.

Keywords: Education; Fidelity; Psychometrics; Simulation

Introduction

Starting from medical school and continuing throughout their careers, health care professionals are exposed to a wide variety of assessments. These target the evaluation of knowledge, clinical skills, and/or attitudes. Simulations are used in the health professions to assess aspects of clinical competence. They have been integrated into many high-stakes exams including and not limited to the United States Medical Licensing Exam (USMLE), Medical Council of Canada, and Royal College of Physicians and Surgeons of Canada.1

* Corresponding address: Assistant Dean for Medical Education, Chairperson of Curriculum Development, Assistant Professor of Medical Education, College of Medicine, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia.

E-mail: fmunshi@kfmc.med.sa (F. Munshi)

Peer review under responsibility of Taibah University.

© 2015 The Authors.

Production and hosting by Elsevier Ltd on behalf of Taibah University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Simulations are defined as “approximations to reality that require trainees to react to problems or conditions as they would under genuine circumstances.”

A simulation, whether it involves standardized patients (SPs), computerized case management scenarios, mannequins, clinical vignettes, or a combination of these methods, holds great promise for both low-stakes tests in medical schools and for high-stakes licensing and certification assessment. The objective of this literature review is to compare the effectiveness of low- and high-fidelity simulations in teaching and assessing clinical skills.

**Assessment using simulations**

The authenticity of assessing clinical competence has a high priority in evaluating outcomes of learning. This has led to the development of a wide variety of simulation-based assessment instruments. Schuwirth and van der Vleuten categorized simulation-based assessment methods using Miller’s pyramid model of medical competence. Miller’s pyramid has four levels of competence: knows, knows how, shows how, and does.

Paper- and computer-based simulations can test at the levels of knows and knows how. Examples of these methods include the Patient Management Problem (PMP), clinical vignettes followed by multiple choice questions or short answers, extended matching questions, and script concordance tests. Assessments with mannequins or simulated patients (SP) test at the shows how level. Performance in actual health care practice tests at the level of does. No single simulation-based assessment method will assess the entire range of medical competencies. A suitable combination is needed to cover all layers of Miller’s pyramid.

**Simulation fidelity continuum**

The degree of realism or authenticity ranges along a scale from completely artificial to an actual real-life situation. A stem of a patient description or a clinical vignette that entails the examinee to make a clinical decision is a simulation at the low end of the fidelity continuum. Assessments using SPs are at the other end of the fidelity continuum, giving a more realistic context for measuring clinical skills and competencies.

Norcini and McKinley argue that test developers may attempt to recreate actual life situations or elaborate tasks in a simulation that will result in a long test with narrow sampling. Due to practical constraints, a long test will contain few problems, which limits the generalizability of scores to the domain of interest. They recommend balancing fidelity and breadth of sampling as this will affect reliability, validity, educational impact, feasibility, and acceptability of the assessment method.

The level of fidelity should be appropriate to the type of task and training stage. A novice can achieve similar or higher skills transfer with a simple simulator, for example, a clinical vignette, than with a complex training aid such as a simulated environment. At more advanced levels of training, the level of fidelity should support higher levels of speed and practice of a task. A simulator is best utilized if used in alignment with educational goals that underpin its use within a program.

The effect of high- and low-fidelity clinical simulations on teaching

Many studies have investigated the educational value of simulations and found them to be valuable. Few articles have compared the educational outcomes between high- and low-fidelity simulations. Various disciplines and clinical skills were used to compare the fidelity effect on learning.

From a historical background, the basis for high-fidelity simulations is ascribed toThorndike’s concept regarding the environment and context of learning and application. However, the studies comparing high- and low-fidelity simulators do not totally support this notion.

Matsumoto et al., Lee et al., and de Giovanni et al. found no differences in performance when subjects were trained on a low- or high-fidelity simulator with the skills being assessed. Scerbo et al. found better performance with the group trained on a low-fidelity simulator. By contrast, Crofts et al., Grady et al., and Rodgers et al. found superior performance with higher fidelity simulator training.

The three articles that concluded superior performance with the group trained by a high-fidelity simulator all evaluated performance of both low- and high-fidelity simulator trained groups on the high fidelity simulator. This may bias the findings because the training was carried out on the lower fidelity simulator while the high-fidelity simulator was used for performance evaluation.

In two studies, performance was evaluated with a neutral task because the optimal goal was transfer of knowledge/skills to a real patient. In the de Giovanni et al. study, diagnostic accuracy and communication skills were equivalent regardless of the degree of simulator fidelity. In the Scerbo et al. study, transfer of phlebotomy training to a real patient was better with the low-fidelity trained group. This can be attributed to what Smallman and St. John described as “naïve realism”. This term describes the desire among users for higher fidelity despite contrary evidence regarding its efficacy.

Simulators are meant to support learning objectives. Issenberg et al. reviewed 109 articles for conditions of high-fidelity simulations that lead to effective learning. The authors reported that the 3 conditions that lead to effective learning most cited in 25%–47% of the articles are provision of feedback, repetitive practice, and curriculum integration. These are features of learning, not characteristics of simulation fidelity. The appropriate level of fidelity is dependent on the intended learning goals and cost. Different levels of fidelity may be needed for different objectives and levels of trainees. To facilitate learning and improve user performance of simulations, Smallman and St. John suggest that training systems should be created from a minimalist perspective, presenting only the essential material needed for a given level of performance.

In laparoscopy training, the low-cost and low-fidelity training box for assessment of skills was found to be superior to high-cost and high fidelity virtual reality laparoscopy training. This has also been confirmed in a randomized crossover study. Similar findings were noted in the learning outcomes for a neonatal resuscitation program (NRP). In a randomized controlled study, use of the low- and high-fidelity mannequin simulators resulted in similar
levels of objectively measured NRP outcomes for integrated skills station and teamwork performance.27 On the other hand, high fidelity was more advantageous than low-fidelity in shoulder dystocia training and nursing learning. 15,24 These variations in findings are merely due to the approach of evaluating a simulation-based learning program and the design of the studies. A more comprehensive approach is proposed in the next section.

Approach to simulator evaluation

Norman and his colleagues defined five characteristics for simulations: fidelity, reliability, validity, learning, and feasibility.25 First, fidelity is the extent to which the appearance and behaviour of the simulated system matches the appearance and behaviour of the actual system.7 In 1954, Miller made an important distinction in simulations between the engineering or physical fidelity and the psychological or functional fidelity. Engineering fidelity is the extent to which the simulation replicates the actual physical characteristics of the actual task. This involves the environment and simulation device or tool. Psychological fidelity is the extent to which the skills of the real task are captured by the simulated task.26 Matthews and Yachmetz described four levels of simulation fidelity. The aim was to develop a common language in clinical simulation terminology.27

Level one (SF1) is state of the art fidelity, in which a computer-aided mannequin interacts with the trainees. Level two (SF2) is high fidelity, which includes a complex scenario that may involve multiple mannequins. Level three (SF3) is intermediate fidelity, which involves a series of procedures put together to resemble a simple scenario found in a clinical setting. Level four (SF4), low fidelity, is meant to demonstrate a simple skill; for example, airway management on an intubation head.

In contrast, Ebel and Frisbie operationally defined reliability as the following: “The reliability coefficient for a set of scores from a group of examinees is the coefficient of correlation between that set of scores and another set of scores on an equivalent test obtained independently from the members of the same group”.28 The definition has three main points that should be highlighted. First, reliability is a measure of the measurement tool when applied to a specific group of participants. The better the fit between the exam difficulty and participant ability, the higher the reliability. Second, the definition states that a correlation coefficient is used as a measure of reliability. A property of the correlation coefficient is that it is a relative not an absolute measure. This implies that if the differences between the scores for the same candidate are small relative to the differences between the scores for different candidates, the test will yield high reliability.28 The definition emphasizes two or more independent measures obtained from equivalent tests of the same construct for each candidate within the group being assessed. This mandates the foundation of the various types of reliability such as test-retest, equivalent forms, and split-halves.

Third, validity is defined as the extent to which a measurement tool measures the intended outcome. Palter and Grantcharov categorized the tools for assessing technical skills in clinical simulations into high and moderate validity.29 High-validity tools include global rating scales and virtual reality simulators. The global rating scale relies on the assessment of an expert observer using predefined criteria to assess ability. Virtual reality simulators provide immediate feedback on performance by recording performance metrics such as time. Moderate-validity tools include checklist scales, analysis of dexterity, and procedure-specific checklists. Checklist scales and procedure-specific checklists are observational tools that rely on an expert assessor. Analysis of dexterity measures performance by the speed and number of a surgeon’s hand movements. The evidence for validity of all of the high- and moderate-validity measurement tools were from observational studies.29

Fourth, the learning impact of simulation studies can be categorized by Kirkpatrick’s four levels of training evaluation: reaction, learning, behaviour, and results.30 The first level includes the assessment of participants’ reactions to the use of simulation. The second level involves a quantitative indicator to measure the increase in knowledge or improvement of skills due to the learning experience. The third level is the assessment of the behavioural change or knowledge application in improving performance in a clinical setting. In other words, the third level assesses the transfer of skills learned through simulations to actual clinical encounters. Level four involves a broader impact of simulation training on the organization.

Last, feasibility describes the degree to which the simulation is affordable and feasible to implement.25 The two criteria to judge feasibility are required cost and value to be attained.

Conclusion

Fidelity plays an important role in the choice of an appropriate simulation for a specific task. High fidelity is not always superior to lower-fidelity because this is dependent on the type of task and the learner’s level. The comparisons made between high- and low-fidelity simulations mainly investigated the educational impact. The psychometric advantages and disadvantages were evidently not elaborated. A more comprehensive evaluation of simulation training should include fidelity, reliability, validity, impact on learning and feasibility. Such an approach can be adopted to compare the psychometric advantages and disadvantages of high- and low-fidelity simulations in future studies.

Conflict of interest

The authors have no conflict of interest to declare.

References


10. Thorndike EL. *Educational psychology.* Lemke and Buechner; 1903.


