



Case-Based Teaching: Does the Addition of High-Fidelity Simulation Make a Difference in Medical Students' Clinical Reasoning Skills?

Mary Kathryn Mutter¹ · James R. Martindale² · Neeral Shah³ · Maryellen E. Gusic² · Stephen J. Wolf⁴

Published online: 10 January 2020

© International Association of Medical Science Educators 2020

Abstract

Context Situativity theory posits that learning and the development of clinical reasoning skills are grounded in context. In case-based teaching, this context comes from recreating the clinical environment, through emulation, as with manikins, or description. In this study, we sought to understand the difference in student clinical reasoning abilities after facilitated patient case scenarios with or without a manikin.

Methods Fourth-year medical students in an internship readiness course were randomized into patient case scenarios without manikin (control group) and with manikin (intervention group) for a chest pain session. The control and intervention groups had identical student-led case progression and faculty debriefing objectives. Clinical reasoning skills were assessed after the session using a 64-question script concordance test (SCT). The test was developed and piloted prior to administration. Hospitalist and emergency medicine faculty responses on the test items served as the expert standard for scoring.

Results Ninety-six students were randomized to case-based sessions with ($n = 48$) or without ($n = 48$) manikin. Ninety students completed the SCT (with manikin $n = 45$, without manikin $n = 45$). A statistically significant mean difference on test performance between the two groups was found ($t = 3.059$, $df = 88$, $p = .003$), with the manikin group achieving higher SCT scores.

Conclusion Use of a manikin in simulated patient case discussion significantly improves students' clinical reasoning skills, as measured by SCT. These results suggest that using a manikin to simulate a patient scenario situates learning, thereby enhancing skill development.

Keywords Simulation-based medical education · Case-based teaching · Undergraduate medical education · Clinical reasoning · Script concordance testing

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40670-019-00904-0>) contains supplementary material, which is available to authorized users.

✉ Mary Kathryn Mutter
mkw2y@hscmail.mcc.virginia.edu

Stephen J. Wolf
<https://orcid.org/0000-0003-0398-9785>

¹ Department of Emergency Medicine, University of Virginia School of Medicine, PO Box 800699, Charlottesville, VA 22908-0699, USA

² University of Virginia School of Medicine, Box 800866 Suite 2008 Pinn Hall, Charlottesville, VA 22908-0866, USA

³ Division of Gastroenterology and Hepatology, University of Virginia School of Medicine, Box 800708, GI West Complex - 2nd Floor, Charlottesville, VA 22908, USA

⁴ Department of Emergency Medicine, Denver Health Medical Center, University of Colorado School of Medicine, Denver Health Medical Center, 777 Bannock St., Pavilion A, Denver, CO 80204, USA

Introduction

In 1999, the Accreditation Council for Graduate Medical Education and the American Board of Medical Specialties defined six domains of competence that all practicing physicians should possess [1]. The Physician Competency Reference sets further outlines each of these domains and delineates how clinical reasoning, defined as skills in diagnosing and managing a clinical problem, contributes to competence in providing patient-centered care [2]. New interns are often on the frontline of patient management and must be able to recognize acute illness and initiate diagnostic and therapeutic decision-making. In order to ensure that trainees are prepared for practice on day one of residency, medical students must develop a strong foundation in clinical reasoning, including for urgent and high acuity scenarios. Residency program directors have emphasized clinical reasoning and recognition of ill patients as important skills for fourth years to acquire

before internship, and several of the Core Entrustable Professional Activities for Entering Residency (CEPAERs) highlight the importance of a graduating student's ability to make informed diagnostic and therapeutic decisions [3–5]. EPA 10, the ability to recognize a patient requiring urgent or emergent care and initiate evaluation and management, specifically outlines how clinical reasoning skills contribute to the performance of this task. Despite the expectation that students should graduate ready to apply sound clinical reasoning and perform all of the core EPA tasks with indirect supervision, in acute situations, students have not met expected levels of performance [6–8].

Situativity theory suggests that the environmental context of learning significantly impacts the development of clinical reasoning skills [9, 10]. Preparing students to apply clinical reasoning skills in acute care situations, however, presents a challenge in undergraduate medical education. Given the unpredictability of acute patient situations, as well as concerns about patient safety, acute care teaching often relies on classroom learning. Case-based teaching can use various approaches to “role-play” details of the case and reveal information as the case progresses, yet lacks “real-world” situated context. Simulation-based medical education (SBME) and simulation-based learning (SBL) are terms used to describe a case-based teaching method involving manikin simulators. High-fidelity simulators are intended to mimic closely patients' physiologic response within clinical care, but use of these tools is a resource intensive endeavor. Recreating the clinical environment incorporates the use of a plastic manikin; equipment, such as cardiac monitors, intravenous lines, and oxygen tubing; and staff, such as simulation technicians and health professions faculty. The application of situativity theory would suggest that the environmental factors present in a SBME learning session, such as the sounds of a cardiac monitor or patient voicing distress, help to simulate the time-pressure of an acute clinical scenario. These contextual factors have a distinct impact on learning and the development of reasoning skills [11].

When comparing SBME to non-manikin case-based teaching, studies have demonstrated that SBME increases learners' self-reported satisfaction with the experience and leads to improvements in knowledge (assessed on multiple choice tests) [12–15]. Studies using oral examinations or a simulation checklist to measure learners' skills in assessing and managing patient after SBME learning have had mixed results [16–23]. While clinical reasoning underlies patient assessment and management skills, these previous studies have not explicitly examined the effect of a manikin on a learner's clinical reasoning skills in areas of high uncertainty. Given the expense of SBME compared with other types of case-based teaching, targeted use of simulation is critical [24].

Script concordance testing (SCT) is an assessment tool to measure clinical reasoning skills using scenarios in which

there is clinical uncertainty [25–27]. SCT is a written test using short clinical scenarios followed by additional clinical data. The learner determines whether the data presented makes a specific diagnosis more or less likely or a management decision (investigation or therapy) indicated or not. The answer is formatted as a 5-point Likert scale (–2, –1, 0, +1, 1). An example of the question format is seen in Fig. 1.

In this randomized controlled trial, we investigated the impact of manikin use on students' clinical reasoning skills in an acute illness patient case scenario, measured by performance on a script concordance test (SCT). We hypothesized that the contextual environment and experience of learning with the manikin would lead to higher clinical reasoning scores on a standardized examination.

Methods

Overall Design

During March 2018, students were randomized to a patient case scenario with manikin (CS-M, control group) or a patient case scenario without manikin (CS-NM, intervention group). Both groups engaged in learning exercises associated with six chest pain case scenarios during the session. All students took a 64-question script concordance test and self-assessment survey after completion of the session. See Fig. 2 for study design. The study was reviewed by the University of Virginia Institutional Review Board for the Social and Behavioral Sciences and was determined to be exempt from further review.

Setting and Participants

Fourth-year medical students at the University of Virginia School of Medicine are required to enroll in an internship readiness course prior to graduation. The 2-week course provides intensive review and the opportunity to practice skills in active learning sessions. Courses are available in Pediatrics, Internal Medicine/Acute Care, and Obstetrics and Gynecology/Surgery and are designed to prepare students for the start of residency in their chosen specialty.

Ninety-six students were enrolled in the Internal Medicine/Acute Care (IM/AC) course and participated in the chest pain case-based session. Students were randomized to a 2-h case session with six chest pain scenarios in groups of six students: with 48 students assigned to the case session with a manikin (CS-M) and 48 students assigned to a session without a manikin (CS-NM). All students provided written consent prior to participating in the session.

Fig. 1 Example of SCT question format

65 y/o F with a history of CAD and type II diabetes mellitus presents with 5/10 pressure-like left-sided chest pain for 1 day.

If you were considering the usefulness of the following investigation...	...and the following new information were to become available...	...you would then consider the investigation...
Q1 Chest ray	Clear lungs bilaterally	-2 -1 0 +1 +2
Q2 ECG	Hypotension and diaphoresis	-2 -1 0 +1 +2
Q3 Troponin	Pain relief with nitroglycerin	-2 -1 0 +1 +2
Q4 Troponin	Pain with inspiration	-2 -1 0 +1 +2

Investigative option:
 -2: useless
 -1: less useful
 0: neither more nor less useful
 +1: useful
 +2: very useful

The Learning Intervention

All students independently listened to a pre-recorded audio file reviewing how to approach an acutely ill patient with chest pain before the start of the session. The room, content, details related to the clinical progression of the case, student prompts, and debriefing methods were the same for CS-M and CS-NM sessions. The sessions were formatted to isolate manikin use (with associated monitors, intravenous lines, and other equipment) as the only difference between the CS-M and CS-NM sessions. In both session types, a student was assigned to be the “team leader” and two other students participated as teammates (Fig. 3). In both the CS-M and CS-NM groups, the team leader was asked to manage progression through the case, with stimuli to prompt decision-making designated by changes in the status of the manikin-patient (CS-M) or described by the faculty instructor (CS-NM). Participating students in both CS-M and CS-NM groups were able to offer work-up or management suggestions and could be assigned care responsibilities (i.e., manage the airway with a bag-valve mask, place orders for labs). The other three students in the group observed during the case progression and then participated in the debriefing. All students in both groups were able to actively participate in three cases, with one time as a team leader. Debriefing occurred immediately after each case concluded

in both the CS-M and CS-NM groups and included all six students.

The case scenarios used in the session were written by a team of three emergency medicine faculty and were then reviewed by three faculty members of the research team (MKM, SJW, NS). A group of three emergency medicine and one internal medicine faculty with experience in case-based teaching, small group facilitation, and structured debriefing in SBME led all the chest pain sessions (MKM, SJW, NS). The faculty as a group rehearsed the case progression and standardized debriefing with predefined learning objectives for all the cases, in order to assure the students received standardized teaching and case progression regardless of their group.

After completion of the learning session (all six chest pain scenarios), study participants completed a brief online survey to assess their perspectives about their engagement during the session and confidence and preparedness to apply clinical reasoning skills in actual patient scenarios. The questions contained Likert-type response options that ranged from Strongly Disagree to Strongly Agree.

The morning after the session, all of the students completed a written 64-question script concordance test ([electronic supplementary material](#)). The script concordance test items were written by three members of the research team, two emergency medicine physicians and an internal medicine physician (MKM, SJW, NS) using published guidelines for test development [26, 27]. Each item was written and revised collaboratively until consensus was achieved amongst all three members of the team. To pilot test the exam, two emergency medicine and two internal medicine faculty, who were not involved in creating or reviewing the test items, took the exam and gave feedback on questions that were not clear or needed improvement. The test was modified by members of the research team using this feedback. Half the test questions (32) used acute chest pain case stems for patients requiring inpatient care. Half of the questions (32) were related to a mixture of other common urgent or emergent inpatient care scenarios such as dizziness, shortness of breath, or confusion. The final version of the exam was administered to ten internal medicine hospitalist faculty and eleven emergency medicine faculty who were not involved in test development or review. In alignment with published guidelines, no training was provided to

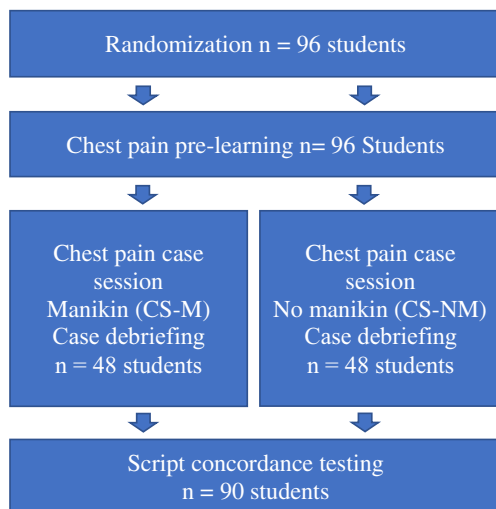
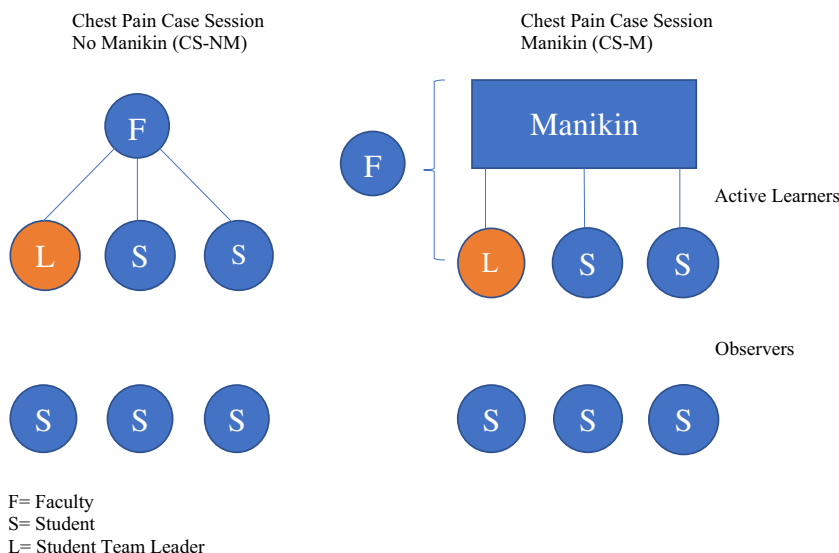


Fig. 2 Study design

Fig. 3 Chest pain session format



these faculty whose scores on the exam were used as the expert standard to measure student performance [26, 27]. Cronbach coefficient alpha was used to determine reliability of the expert standard. The overall Cronbach coefficient alpha score for the expert tests was 0.68, which is slightly lower than the recommended score of 0.75 [26]. Due to time limitations, we administered the test without further modification.

Data Analysis

The scoring of the SCT followed the steps described by Fournier [26]. The points awarded to the students’ tests were calculated using the scores from the panel of faculty test scores. The score for each answer was the number of experts giving a particular answer divided by the modal answer. In Fig. 1, for example, question 1 may have 14 of 20 experts who answered + 1, and 6 of 20 experts who answered 0. Therefore, a student who answered + 1 would get 1 point (14/14), the answer 0 would get 0.4 (6/14), and all other answers on the Likert scale would get 0 points. The final student score was created by adding all question scores together, dividing by 64 (total questions) and multiplying by 100 for a percentage. Although each item on the exam is scored relative to the expert standard for the item, only an overall score representing aggregate performance on all test items is used to present exam results on a SCT test (26).

An independent samples *t*-test was used to compare the mean student scores on the 64-item SCT between the manikin (CS-M) or non-manikin (CS-NM) groups. An alpha level of .05 was used to determine statistical significance. Data were analyzed using SPSS version 25. In order to assess if differences existed between CS-M and CS-NM groups across all five categories of responses on the survey, a Mann-Whitney *U* test was used. A Bonferroni correction was applied in order to

account for the increased Type 1 error rate associated with running multiple tests with the same groups.

Results

Ninety-six students agreed to participate in the study. Ninety SCT examinations were available for analysis. Three examinations were not available for analysis from both the control and the intervention groups (three students were excused from the examination session, two tests were not submitted to course faculty, and one test was submitted but was not labeled to indicate which type of learning session the student attended).

The mean percent correct on the SCT exam for students in the CS-M group was 67.23 (SD = 6.05). The overall mean performance of this group was significantly higher ($t = 3.059$, $df = 88$, $p = .003$, 95%CI 1.34–6.32) than the mean performance of the CS-NM group (mean percent correct 63.39, SD = 5.83) (Table 1).

Forty-six students from the CS-M group and forty-eight students from the CS-NM group provided responses to the survey questions. Each of these students responded to all four questions on the survey; there were no students who responded to only some of the survey questions. Although the number and percentage of students in each group who selected specific responses on the Likert scale differed, there were no statistically significant differences between the two

Table 1 SCT exam scores for CS-M vs CS-NM groups

Group	<i>N</i>	Mean	Std. deviation
Score			
Manikin (CS-M)	45	67.23	6.05
No Manikin (CS-NM)	45	63.40	5.84

groups on any of the questions on the post-course survey (Table 2). For example, in response to the prompt, “I am more confident in my ability to make decisions in urgent clinical situations,” 21 students in the CS-M group strongly agreed, while 11 students in the CS-NM group strongly agreed. For this prompt, 39 CS-M students and 43 CS-NM students agreed or strongly agreed.

Discussion

Learners reported high levels of engagement in both the manikin and non-manikin sessions. It is not clear what factors contribute to an individual student’s engagement in learning, and it is likely that unique aspects of an experience make a student feel engaged. Aligned with Kirkpatrick’s framework for evaluating educational programs, measurements of participants’ reactions, including engagement, is a lower level outcome of the impact of the intervention; however, it is unlikely that higher level outcomes (change in knowledge, skills, and behaviors) will occur if a learner is not engaged [28].

Previous studies have demonstrated higher measures of knowledge, skills, and decision-making when learners engage in manikin-based learning sessions. In these studies, assessment and decision-making skills were assessed by observers using performance checklists. Information related to observer training and interrater reliability were not included in all studies, raising concerns about the potential impact of observer bias [29].

Script concordance testing has been shown to be a useful tool to measure clinical reasoning for learners across the continuum of education [30–33]. SCT is an objective assessment tool in which authentic clinical scenarios can be used to measure reasoning skills in situations where the next best step in diagnosis or management is uncertain. In this study, we assessed students’ clinical reasoning skills after participating

in a simulated patient case session, comparing performance on a SCT for students whose session included use of a manikin (CS-M) with performance of students who engaged in the same format session without a manikin (CS-NM). Students who participated in the CS-M session achieved scores that were significantly higher than students in the CS-NM session, with a difference of 3.84. This SCT score increase is consistent with score increases between post-graduate year (PGY) 1 residents and PGY-3’s, in a previous study of lumbar puncture clinical decision-making in pediatric residents [30]. This finding supports our hypothesis that the contextual environment provided by the manikin improves learning and promotes the development of clinical reasoning skills in cases involving acute patient care scenarios. Additionally, our study strengthens the existing research that supports the use of SBME for skill acquisition in acute clinical care scenarios [24, 34, 35].

Ultimately, it is important to be able to measure learners’ clinical reasoning in actual patient settings (transfer of learning) and the impact of sound decision-making on patient care outcomes. We used a proxy measure of behavior change in this study: students’ self-reported confidence and feelings of preparedness to apply the skills they learned in future patient scenarios. Feelings of self-efficacy are felt to correlate with increased likelihood of success in applying the skills one has learned [36–38]. In our study, students had the opportunity to reflect on the learning experience (CS-M and CS-NM, each with facilitated debriefing) before responding to questions about their confidence and preparedness. We hypothesized that the hands-on experience and environment of the CS-M session would lead to higher levels of self-efficacy, but students in both the CS-NM and CS-M groups expressed feeling very confident in their abilities and well prepared for

Table 2 Self-assessment survey results with statistical comparison between CS-M and CS-NM groups

	Group	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Standardized Mann-Whitney U (Z)	Sig.(2-tailed)
I was engaged throughout the session	CS-M	0	0	0	14 (30.4%)	32 (69.6%)	-.83	.41
	CS-NM	3(6.3%)	0	2 (4.2%)	12 (25.0%)	31 (64.6%)		
I am more confident in my ability to make decisions in urgent clinical situations	CS-M	0	0	7 (15.2%)	18 (39.1%)	21 (45.7%)	-1.55	.12
	CS-NM	0	0	5 (10.4%)	32 (66.7%)	11 (22.9%)		
I am more confident in my ability to make clinical decisions independently	CS-M	0	1 (2.2%)	8 (17.4%)	23 (50.0%)	14 (30.4%)	-1.12	.27
	CS-NM	1 (2.1%)	1 (2.1%)	9 (18.8%)	28 (58.3%)	9 (18.8%)		
This session helped to prepare me for internship	CS-M	0	0	4(8.%)	20 (43.5%)	22 (47.8%)	-.60	.55
	CS-NM	0	1(2.1%)	2(4.2%)	26(54.2%)	19(39.6%)		

future clinical experiences. Standardization of content and debriefing sessions, as well as use of highly experienced faculty instructors, likely contributed to high levels of student confidence.

Our findings have significant implications for educators interested in designing learning experiences to aid students in developing clinical reasoning skills. Importantly, to prepare students to transition to graduate medical education and ensure that they are ready to manage acute clinical scenarios on day one of internship, it is critical to consider how the context and experience of a learning session will contribute to and promote learning. Internship readiness or “bootcamp” courses should use various pedagogies; SBME, in particular, should be used with intentionality to make the best use of limited or expensive resources and provide effective and efficient educational sessions. Our results support the use of SBME in courses near the transition from undergraduate to graduate medical education, in order to help students develop the clinical reasoning skills they will need in urgent and emergent patient care settings.

Limitations

This is a single center study and we did not use a before and after measure of clinical reasoning. We did not address EPA 10 specifically, but rather a fundamental skill, clinical reasoning in acute scenarios, that underlies the ability to perform this task. Although students in this study were exposed to assessment of EPAs related to history taking, physical examination, ability to generate a differential diagnosis, and to perform an oral presentation and document encounters in written notes, in three rotations during their clerkship year, our curriculum had not implemented teaching or assessment of EPA 10 at the time of the study. Students were randomized in order to mitigate potential contribution of a priori differences in the groups. All of the students who participated in the study were students exposed to the same curricular opportunities at the School of Medicine, including a course focused on clinical skill development in the pre-clerkship phase of the curriculum and a transition to clerkship course. Despite efforts to standardize the teaching for the chest pain sessions in the readiness course, four different faculty taught the sessions and variability in teaching style or skills could have contributed to the results of the study. We attempted to minimize any variability by providing pre-session practice and standardization exercises for the faculty and by assigning two faculty to run all the CS-M sessions and two faculty to run all the CS-NM sessions.

Finally, the SCT exam was created, reviewed, and pilot tested by faculty at the institution with clinical expertise in emergency medicine and internal medicine prior to administration to the faculty whose scores served as the expert standard. To ensure reliability of the expert standard, we measured Cronbach coefficient alpha. Our result was slightly lower than

recommended level to measure reliability of a test (26). This could have affected our ability to accurately discriminate student performance.

Conclusion

In a fourth-year internship readiness course, students participating in simulated patient case sessions with a manikin demonstrated significantly higher scores on a test of clinical reasoning in acute care scenarios. This study offers further evidence that SBME is an effective strategy to use in transition to residency courses, particularly when learning objectives include the development of clinical reasoning in patient care scenarios that students are less likely to have encountered in authentic clinical experiences. Use of a manikin augments a simulated experience that will prepare learners to apply clinical reasoning skills as interns caring for patients in urgent and emergent settings.

Acknowledgments We thank Dr. Meredith Thompson for her significant contribution to this project.

Authors' Contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by M. Kathryn Mutter, James R. Martindale, Neeral Shah, and Stephen J. Wolf. Data analysis was performed by James R. Martindale and M. Kathryn Mutter. Supervision was provided by James R. Martindale, Stephen J. Wolf, and Maryellen E. Gusic. The first draft of the manuscript was written by M. Kathryn Mutter, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This project was supported by an internal Educational Fellowship Award at the University of Virginia School of Medicine.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This research project was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The University of Virginia Institutional Review Board for Social and Behavioral Sciences declared the study exempt (Reference number 2018006700).

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)*. 2002;21(5):103–11.

2. Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrenner CA. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. *Acad Med.* 2013;88(8):1088–94.
3. Lyss-Lerman P, Teherani A, Aagaard E, Loeser H, Cooke M, Harper GM. What training is needed in the fourth year of medical school? Views of residency program directors. *Acad Med.* 2009;84(7):823–9.
4. Angus S, Vu TR, Halvorsen AJ, Aiyer M, McKown K, Chmielewski AF, et al. What skills should new internal medicine interns have in July? A national survey of internal medicine residency program directors. *Acad Med.* 2014 Mar;89(3):432–5.
5. Englander R, Flynn T, Call S, et al. Core entrustable professional activities for entering residency: Curriculum developers guide. Washington: Association of American Medical Colleges MedEdPORTAL iCollaborative; 2014. Available at: <https://icollaborative.aamc.org/resource/887/>. Accessed 12/3, 2018
6. McEvoy MD, Dewaay DJ, Vanderbilt A, Alexander LA, Stillely MC, Hege MC, et al. Are fourth-year medical students as prepared to manage unstable patients as they are to manage stable patients? *Acad Med.* 2014;89(4):618–24.
7. Young JS, Dubose JE, Hedrick TL, Conaway MR, Nolley B. The use of “war games” to evaluate performance of students and residents in basic clinical scenarios: a disturbing analysis. *J Trauma.* 2007;63(3):556–64.
8. Fessler HE. Undergraduate medical education in critical care. *Crit Care Med.* 2012;40(11):3065–9.
9. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educ Res.* 1989;18(1):32–42.
10. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach.* 2011;33(3):188–99.
11. Patel R, Sandars J, Carr S. Clinical diagnostic decision-making in real life contexts: a trans-theoretical approach for teaching: AMEE Guide No. 95. *Med Teach.* 2015;37(3):211–27.
12. Ten Eyck RP, Tews M, Ballester JM. Improved medical student satisfaction and test performance with a simulation-based emergency medicine curriculum: a randomized controlled trial. *Ann Emerg Med.* 2009;54(5):684–91.
13. Smithburger PL, Kane-Gill SL, Ruby CM, Seybert AL. Comparing effectiveness of 3 learning strategies: simulation-based learning, problem-based learning, and standardized patients. *Simul Healthc.* 2012;7(3):141–6.
14. Cleave-Hogg D, Morgan PJ. Experiential learning in an anaesthesia simulation Centre: analysis of students’ comments. *Med Teach.* 2002;24(1):23–6.
15. Gordon JA. The human patient simulator: acceptance and efficacy as a teaching tool for students. The Medical Readiness Trainer Team. *Acad Med.* 2000;75(5):522.
16. Steadman RH, Coates WC, Huang YM, Matevosian R, Larmon BR, McCullough L, et al. Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Crit Care Med.* 2006;34(1):151–7.
17. Littlewood KE, Shilling AM, Stemland CJ, Wright EB, Kirk MA. High-fidelity simulation is superior to case-based discussion in teaching the management of shock. *Med Teach.* 2013;35(3):e1003–10.
18. Ten Eyck RP, Tews M, Ballester JM, Hamilton GC. Improved fourth-year medical student clinical decision-making performance as a resuscitation team leader after a simulation-based curriculum. *Simul Healthc.* 2010 Jun;5(3):139–45.
19. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med.* 2011;86(6):706–11.
20. Lorello GR, Cook DA, Johnson RL, Brydges R. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. *Br J Anaesth.* 2014;112(2):231–45.
21. Schwartz LR, Fernandez R, Kouyoumjian SR, Jones KA, Compton S. A randomized comparison trial of case-based learning versus human patient simulation in medical student education. *Acad Emerg Med.* 2007;14(2):130–7.
22. Morgan PJ, Cleave-Hogg D, McLroy J, Devitt JH. Simulation technology: a comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology.* 2002;96(1):10–6.
23. Wenk M, Waurick R, Schotes D, Wenk M, Gerdes C, Van Aken HK, et al. Simulation-based medical education is no better than problem-based discussions and induces misjudgment in self-assessment. *Adv Health Sci Educ Theory Pract.* 2009;14(2):159–71.
24. Cook DA, Brydges R, Hamstra SJ, Zendejas B, Szostek JH, Wang AT, et al. Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. *Simul Healthc.* 2012;7(5):308–20.
25. Charlin B, Brailovsky C, Leduc C, Blouin D. The diagnosis script questionnaire: a new tool to assess a specific dimension of clinical competence. *Adv Health Sci Educ Theory Pract.* 1998;3(1):51–8.
26. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak.* 2008;8:18–6947 -8-18.
27. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach.* 2013;35(3):184–93.
28. Moreau KA. Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Med Teach.* 2017;39(9):999–1001.
29. Mahtani K, Spencer EA, Brassey J, Heneghan C. Catalogue of bias: observer bias. *BMJ Evid Based Med.* 2018;23(1):23–4.
30. Chang TP, Kessler D, McAninch B, Fein DM, Scherzer DJ, Seelbach E, et al. Script concordance testing: assessing residents’ clinical decision-making skills for infant lumbar punctures. *Acad Med.* 2014;89(1):128–35.
31. Carriere B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a script concordance test. *Ann Emerg Med.* 2009;53(5):647–52.
32. Humbert AJ, Besinger B, Miech EJ. Assessing clinical reasoning skills in scenarios of uncertainty: convergent validity for a Script Concordance Test in an emergency medicine clerkship and residency. *Acad Emerg Med.* 2011;18(6):627–34.
33. Goulet F, Jacques A, Gagnon R, Charlin B, Shabah A. Poorly performing physicians: does the Script Concordance Test detect bad clinical reasoning? *J Contin Educ Heal Prof.* 2010;30(3):161–6.
34. Beal MD, Kinnear J, Anderson CR, Martin TD, Wamboldt R, Hooper L. The effectiveness of medical simulation in teaching medical students critical care medicine: a systematic review and meta-analysis. *Simul Healthc.* 2017;12(2):104–16.
35. McGaghie WC, Issenberg SB, Petrusa ER, Scales RJ. Revisiting ‘A critical review of simulation-based medical education research: 2003-2009’. *Med Educ.* 2016;50(10):986–91.
36. Bandura A. Self-efficacy mechanism in human agency. *Am Psychol.* 1982;37:122–47.
37. Eva KW, Cunningham JP, Reiter HI, Keane DR, Norman GR. How can I know what I don’t know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract.* 2004;9(3):211–24.
38. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005 Oct;80(10 Suppl):S46–54.