Learning Empathy Through Simulation

A Systematic Literature Review

Margaret Bearman, PhD;
Claire Palermo, PhD;
Louise M. Allen, B NutrDiet (hons);
Brett Williams, PhD

Summary Statement: Simulation is increasingly used as an educational methodology for teaching empathy to preservice health professional students. This systematic review aimed to determine if and how simulation, including games, simulated patients, and role-play, might develop empathy and empathetic behaviors in learners. Eleven databases or clearing houses including MEDLINE, EMBASE, CINAHL, PsychInfo, and ERIC were searched for all articles published from any date until May 2014, using terms relating to (i) preservice health professional students, (ii) simulation, and (iii) empathy. Twenty-seven studies met the inclusion criteria, including 9 randomized controlled trials. A narrative synthesis suggests that simulation may be an appropriate method to teach empathy to preservice health professional students and identifies the value of the learner taking the role of the patient.
(*Sim Healthcare* 10:308–319, 2015)

Key Words: Simulation, Empathy, Role-play, Communication skills, Simulated patients.

Simulation-based education, where students engage in an experience to learn, is frequently used to develop empathy and empathetic behaviors in medical,[1] nursing,[2] and allied health[3] students. Simulation techniques in health professional education includes a variety of different approaches such as simulated or standardized patient (SP) methodology, mannequin-based methodologies, role-play, games, and virtual reality.[4] This approach is not universally accepted. Wear and Varley[5] caution that simulated empathy lacks authenticity as students learn to act "empathetic" for purposes of performance rather than establishing a genuine connection with real people, a notion described as learning to play the "simulation game."[6]

The value of empathy to health care practice is not controversial. Hojat[7] distinguishes empathy, with its predominantly cognitive and altruistic orientation, from sympathy, which is denoted as a predominantly emotional and self-serving orientation. This distinction, which has its critics,[7] is maintained in the associated definition of empathy in health care as a cognitive response of understanding "the experiences, concerns, and perspectives"[7] of the patient, including the capacity to communicate this understanding. This emphasis on the behavioral manifestations of empathy is notable as the latter is easier to objectively measure. Some measures of empathy do emphasize the internal experience of empathy,[8] and others again require the perspective of a patient.[9]

This intuitive notion that practitioners' empathetic behaviors improve the delivery of care is increasingly supported by evidence, such as the better outcome of diabetic patients associated with more empathetic doctors[10,11] and the reduction of seclusion and restraint of psychiatric patients associated with more empathetic nurses.[12] However, the value of health professional education in teaching empathy is not so clear. On the one hand, the 2013 systematic review conducted by Batt-Rawden et al[13] of 18 educational interventions to teach empathy to medical students from 2003 to 2012 concluded that the interventions were mostly effective in promoting empathy despite methodological flaws. This review identified the success of some approaches to teaching empathy, which might be considered simulation, such as "experiential learning" and "drama." Other studies[14,15] also indicate the value of empathy education for practitioners. On the other hand, a recent systematic review found that empathy seems to decline during medical education and residency.[16] Similar findings can be found in other disciplines.[17] Neumann et al[16] believe that this may be a consequence of entering into the clinical environment, which increases feelings of vulnerability; they also note the role of idealistic belief in the role of doctor and some students' generally heightened levels of distress (eg, burnout, depression).

This article builds on previous reviews by specifically focusing on simulation as the learning strategy. It aimed to determine if simulation-based education of preservice health care professionals, in comparison with an alternative or no intervention, is associated with improved empathetic behaviors. The secondary aim was to explore the key learning and teaching approaches, if any, that are associated with the improved outcomes.

METHODS

Literature Search and Study Selection

Eleven databases or clearing houses (MEDLINE, EMBASE, CINAHL, PsychINFO, ERIC, Web of Science, Scopus, Informit, Campbell Collaboration, BEME, and Cochrane) were searched for all records from any date up to May 2, 2014. Search terms

From the HealthPEER (M.L.B.), Department of Nutrition and Dietetics (C.P., L.M.A.), and Department of Community Emergency Health and Paramedic Practice (B.W.), Monash University, Victoria, Australia.

Reprints: Margaret Bearman, PhD, HealthPEER, G08 13c Monash University, Clayton Vic 3800, Australia (e-mail: argaret.bearman@monash.edu).

DOI: 10.1097/SIH.0000000000000113

were grouped into 3 conceptual categories as follows: (i) health professional students, (ii) simulation, and (iii) empathy. Search terms regarding health professional education students included variants of 18 professions combined with variants of *education*, *learner*, *student*, and *teaching*. Search terms regarding simulation included variants on *clinical skill*, *interactive computer*, *fish bowl*, *haptic*, *manikin/mannequin*, *role-play*, *simulated/standardized patient*, *virtual environment*, and *simulation*. Search terms regarding empathy included *compassion*, *communication skills*, *doctor-patient relations*, *emotions*, *patient-centred communication*, and *student-patient communication*. A specialist librarian oversaw the development and implementation of the search strategies. The initial yield of studies for review contained studies that fulfilled all 3 search concepts, that is, articles that contained search terms for health professional students AND simulation AND empathy. The following limitations were applied: English language, peer reviewed, and availability of the full-text article.

Duplicates and studies that were unrelated to the review question were excluded on title. Ten articles were selected from the yield to pilot the inclusion/exclusion criteria. All authors rated all 10 articles and finalized criteria through consensus. The final criteria for inclusion of a publication were as follows: first, the population of the study was preservice (prelicensure) health professional students; second, the simulations could be any modality including role-play, simulated patients, virtual, or mannequin based; third, that there was some comparison to assess improvement in empathetic behaviors; and finally, that there be some kind of qualitative or quantitative assessment of empathy. In the last case, we included studies that contained subscales or even single items, if the intention was to assess empathy. Veterinary students, cases where there was no enactment of a situation (eg, part-task trainers, paper cases), commentaries or similar, studies that investigated communication without specific mention of empathy, or experiences that were not educational were excluded. Systematic reviews that matched the inclusion criteria were acceptable. All remaining abstracts were then independently examined by 2 of the authors against the inclusion/exclusion criteria for progression to full-text articles.

All remaining studies were read in full text by 2 of the authors and independently examined against the inclusion/exclusion criteria. Agreement was reached through discussion and negotiation.

Data Extraction

Data were extracted from all eligible articles, including the location of study, student sample, study design, simulation modality, comparator, outcomes, and results. Data were extracted by 2 of the authors, and agreement was reached by negotiation. Two of the authors (M.B. and either C.P. or L.M.A.) finalized terms, gaps, and discordances between reviewers. Where effect sizes were not reported, Cohen *d* was calculated from available data.

Quality Assessment

The included studies presented either entirely or predominantly quantitative data and so were assessed for quality against the Medical Education Research Study Quality Instrument (MERSQI).[18] The MERSQI was designed specifically for quantitative observational, quasi-experimental, and experimental studies in medical education and has been tested with respect to item and rater reliability, principal components, and criterion validity.[18] Two researchers (M.B. and C.P.) independently assessed the quality of all articles (range, 5–18/18). Differences in interpretation were resolved through consensus.

Synthesis

A narrative, descriptive approach was taken across the studies, drawing from the principles of realist review[19] by focusing on "demiregularities" to elucidate potential mechanisms whereby health professional students learn empathy from simulation-based education. Because of limitations of pre-post studies,[20] randomized comparative studies were considered most closely, and effect sizes were displayed on a forest plot. Effect sizes were not pooled in a meta-analysis because this is not recommended when there is a diversity of comparators[21] or outcomes.[22]

RESULTS

Literature Search and Study Inclusion

From the initial search of 11 databases, a total of 14,748 articles were retrieved, dated up to May 2, 2014. After duplicates and title review, 836 abstracts were admitted to the next phase. After the review based on the inclusion/exclusion criteria, 765 were excluded. Full publications for the remaining 71 articles were retrieved, and 44 were excluded after review. An overview of the study inclusion process is shown in Figure 1.

Data Extraction and Quality Assessment

Table 1 summarizes the 27 included studies. Fourteen studies were published from 2000 onward, 8 from 2012 onward. Professional groups were as follows: medicine (18), nursing (4), pharmacy (2), social work (1), dental hygiene (1), and nutrition/dietetics (1). There were 14 pre-post designs, 9 randomized controlled trials (RCTs) or randomized trials, of which 1 contained a 3-way comparison and 5 had quasi-experimental designs. This includes 1 study that reported both an RCT and pre-post design. Three studies reported supplementary qualitative methods.

Of the 9 RCT studies, 4 (44%) reported significant improvements in learners' empathy or empathetic behaviors between those who learned via simulation and those who were given an n = 2 or no n = 2 alternative (quality assessment range, 11–15.5). Of the 9 RCTs, 3 (33%) reported no significant change between those who learned via simulation and those who were given an n = 1 or no n = 2 alternative (quality assessment range, 9.5–12.5). Of the 9 RCTs, 3 (33%) reported significant differences between different approaches to simulation-based education (quality assessment range, 12.5–15.5). Of the 14 pre-post designs, 12 (86%) reported a significant improvement in learners' measures of empathy (quality assessment range, 7.5–12.15). One pre-post study (7%) reported a decrease in empathy (quality assessment score, 10.5), and 1 (7%) did not conduct

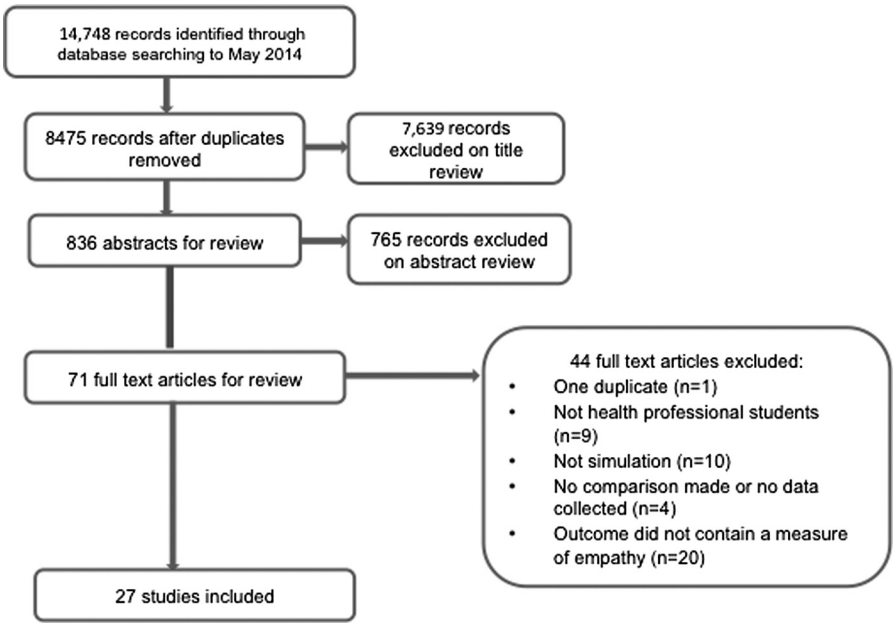

FIGURE 1. Study inclusion process.

statistical analysis (quality assessment score, 8). Of the quasi-experimental designs, 4 (80%) of the 5 reported significant improvement in learners' empathy levels between those who learned via simulation and those who were given an n = 1 or no n = 3 alternative (quality assessment range, 8.5–12.5). Of 5 quasi-experimental studies, 1 (20%) reported no change between those who learned via simulation and those who were given no alternative (quality assessment score, 8.5) (Table 1).

One article reported a pre-post and an RCT study of the same intervention in different years. Cahan et al[21] describe that the pre-post indicated significant learning for a cohort experiencing an intervention when comparative trials showed that there were no significant differences in empathy between simulation group and those who did not have an alternative.

The outcome measures were diverse, and many were subdomains of other scales. Of the 27 studies, 17 (63%) used self-report measures; 8[1–3,22,26,29,40,41] used well-studied attitudinal scales such as the Jefferson Scale of Empathy[36]; 3[25,28,30] used the literature-based but not rigorously studied scale, the "Maxwell and Sullivan" questionnaire; and 6[34,39,42–45] used self-designed questionnaires. Of the 27 studies, 10 (37%) used raters' assessment of behaviors with simulated patients, using previously developed scales or Objective Structured Clinical Examination (OSCE) scores. Raters included simulated patients, trained laypeople, examiners, researchers, and, in one instance, patients. Many studies had a focus on measuring changes in empathetic behaviors through self-ratings or observational ratings; others considered more closely the learner's affect through attitudinal questionnaires. Some studies focused on attitudes to specific patient experiences (obesity, aged care, mental illness). One RCT demonstrated improved empathy after intervention using one different scale, the Arizona Clinical Interview Rating Scale (ACIRS) but not another, the Roter Interactional Analysis System (RIAS).[46]

Synthesis

When the simulation designs of the studies were considered as a whole, with a particular focus on mechanisms that might promote learning, 2 themes were noted as follows:

1. Being a health professional
   These designs required the learner to simulate a health professional. These were generally designed for students to develop skills in communication. The most common designs were interacting with an SP or as part of a peer role-play.
2. Being a patient
   These designs required the learner to simulate the patient. The most common designs were role-play, auditory hallucination simulations, and games about negotiating being elderly.

When the studies are categorized according to these themes, of 27 simulation designs, 10 (37%) required the learner to only simulate the patient, 7 (26%) required the learner to only simulate the health professional, and 10 (37%) (most commonly role-play) required learners to act as both or either patient and health professional. Other groupings, such as type of measure, type of comparator, length of intervention, or content of simulation, did not present as coherent themes.

As noted earlier, one study showed different results between a pre-post study and an RCT. Norman[20] suggests that experimental designs are best for investigating efficacy. He underlines the limitations of pre-post 1-group design, which cannot distinguish improvements that are due to the intervention from "natural" student progression.[20] To reduce this risk of bias and because there were sufficient experimental designs within the included articles, the synthesis process focused on studies where the participants were randomized.

Arms of RCT interventions were categorized according to the 2 themes—*being a patient* and *being a health professional*. Refer to Table 2 for details. As mentioned, there were 3 RCTs[3,21,23] that did not show significant differences in

TABLE 1. Profile of Included Studies

Reference	Location	Sample Size	Study Design	Simulation Modality and Topic	Comparator (Experimental Designs)	Outcome Measure/s (Empathy Only) and Rater	Psychometric Properties of Instrument	Results	MERSQI[18] (Range, 5–18)
Bath et al,[23] 2000	Un	34 final-year medical students (at posttest)	Pre-post	24-h care of a computerized doll simulating 6-wk-old infant	—	Author devised questionnaire: questions regarding empathy toward parents of young children Self-rated	Not reported	3% less empathy 47% unchanged attitudes 35% a little more empathy 15% a lot more empathy Insufficient data to calculate ES	8
Bayne et al,[24] 2001	United States	22 third-year medical students	Pre-post	SP and RP, part of facilitated discussions regarding empathy (2 d)	—	Consultation and Relational Empathy (CARE) Scale Rated by SPs.	CARE referenced in the literature; previous psychometric properties reported as $\alpha = 0.93$ and concurrent validity with other measures, $r = 0.84/r = 0.63$.	Improved empathy ($P < 0.01$, $d = 0.75$)	11.5
Bosse et al,[25] 2012	Germany	103 fifth-year medical students	RCT, 3 group using pre-post measures	SP or RP communication with and counselling of parents of sick children (9 cases)	Established course without RP or SP compared with both SP and RP	Calgary-Cambridge subscale Rated by trained psychologist within an OSCE	Calgary-Cambridge subscale referenced in the literature. α reported as 0.705.	Improved understanding of parents' perspective for RP and SP ($P < 0.001$, d = 1.48) and $P < 0.006$, $d = 0.63$) compared with CG. Higher scores for the RP group than for the SP group ($P < 0.001$, $d = 0.71$)	15.5
Bunn et al,[26] 2009	United States	150 medical students	RCT, 2 group using pre-post measures	40-min simulated auditory hallucination	No intervention	Jefferson Scale of Physician Empathy Student Version (JSPE-S) Self-rated	JSPE-S referenced in the literature.	Improved empathy ($P < 0.0001$) for intervention compared with control ES calculated as $d = 0.60$	11
Cahan et al,[21] 2010	United States	147 third-year medical students	Two study designs: 2-group RCT and pre-post	SP following communication skills training, workshop part of a human factors curriculum	Study 1: no intervention Study 2: no comparator	Author devised assessment rubric grading written response to vignettes Vignettes rated by 1 or 2 authors	Interrater reliability: 288 responses by 2 raters; α reported as 0.808	Study 1. RCT: no significant improvement in empathy ($P < 0.94$) compared with control. ES calculated as $d = 0.02$ Study 2. Pre-post: improved empathy ($P < 0.001$) ES calculated as $d = 1.95$	Study 1, 14; study 2, 12.5
Chaffin and Adams 2013[27]	United States	67 senior nursing students	Pre-post content analysis of written reflections	30- to 40-min simulated auditory hallucination	—	Author devised rating: Likert scale (1–5) regarding self-rated level of empathy Self-rated	None reported	Significantly improved empathy ($P < 0.001$) ES calculated as $d = 3.22$	8.5
Chunharas et al,[28] 2013	Thailand	89 fifth-year medical students	Two groups nonrandomized with postmeasure	Students practice injection skill manikins and on themselves	Students practice injection skill on manikins	Author devised scale (1–4) rating feeling of empathy toward children Self-rated retrospectively assessing empathy at times during the experience	None reported	Intervention group—significantly more students retrospectively rated their empathy as "high" after injecting children as those who retrospectively rated their empathy as high ($P < 0.01$). There was a significant difference between the intervention and control line at the baseline measure of retrospectively rating empathy before injecting children. Insufficient data to calculate ES.	9

(continued on next page)

TABLE 1. *(Continued)*

Reference	Location	Sample Size	Study Design	Simulation Modality and Topic	Comparator (Experimental Designs)	Outcome Measure/s (Empathy Only) and Rater	Psychometric Properties of Instrument	Results	MERSQI[18] (Range, 5–18)
Daeppen et al,[29] 2012	Switzerland	91 fifth-year medical students	RCT	Motivational interviewing after training 2 × 4 h workshops, with significant RP component	No motivational interviewing workshops	Motivational Interviewing Treatment Integrity (MITI) scoring tool empathy subscale. Videos with SPs assessed by trained nonmedical raters	MITI psychometric properties referenced in the literature. ICC = 0.51 reported for empathy subscale ratings.	Empathy subscale higher in the intervention group—($P < 0.001$), ES: $\Theta = 0.73$ (nonparametric)	14.5
Dearing et al,[2] 2006	United States	94 nursing students	1. Two-group nonrandomized comparison 2. Focus groups	45-min simulated auditory hallucination	No intervention	1. Medical Condition Regard Scale (MCRS) Self-rated 2. Themes from focus group transcripts analyzed using schema analysis	MCRS referenced in the literature.	1. There was a significant difference in the posttest MCRS scores of the 2 groups ($P = 0.001$). Significant improvement in 6 items. Analysis was item by item, no ES calculated. 2. Qualitative differences noted between 2 groups in the ability to grasp patients' perspective.	12.5
Deladisma et al,[30] 2007	United States	84 medical students	Two-group randomized trial, single measure of response to intervention	Virtual patient case with abdominal pain	SP	Consensus devised score sheet including observed behavior and specific empathy item. Clinician raters assessed videos for nonverbal communication.	α for empathy measure—reported as 0.92	Higher empathy in the SP group ($P < 0.05$) compared with virtual patient. ES calculated as $d = 0.70$. Empathy positively correlated with eye contact, body lean, head nod, and level of immersion.	14
Dikici et al,[31] 2009	Turkey	146 medical students, level not stated	Pre-post, with 6 mo postrepeated	SP within a breaking bad news sessions (4 cases)	—	Empathetic behavior domain in OSCE score on breaking bad news. Rated as part of OSCE.	None reported	Improved empathetic behaviors immediately after ($P < 0.001$) and 6 mo later ($P < 0.001$). ES calculated pre-post as 14.70.	11
Evans et al,[32] 2005	United States	101 first-year pharmacy students	Pre-post	Geriatric Medication Game (1–1.5 h)	—	Self-devised 8-item questionnaire, regarding attitudes toward the elderly. Self-rated.	None reported	Responses to 6 of 8 items significantly improved after intervention (individual items were compared, significant P ranged from 0.001 to 0.47) Analysis was item by item, no ES calculated.	7.5
Galletly et al,[33] 2011	Australia*	87 final-year medical students	Pre-post	Simulated auditory hallucinations (45 min) as part of session on mental illness	—	Attitudes to Mental Illness Questionnaire (AMIQ) Self-rated	AMIQ referenced in the literature.	Improved attitudes ($P < 0.001$) overall; lowest ½ and lowest ¼ improved significantly but highest ½ and highest ¼ did not. ES calculated as $d = 2.25$.	8.5
Gleber et al,[3] 1995	United States	24 dental hygiene students	RCT, 2-group with pre-post and 12-mo follow-up measures	SP as part of other modalities within 20 h on interpersonal skills training	No intervention	Mehrabian and Epstein Measure of Emotional Empathy Self-rated.	Mehrabian and Epstein Measure of Emotional Empathy referenced in the literature, cited as subscale intercorrelation $r > 0.30$ and significant at 0.01 level of confidence; α cited as 0.84.	No significant improvement in empathy after test ($P < 0.73$) compared with control group or at follow-up. ES calculated as $d = 0.11$.	9.5

Gockel and Burton 2014[34]	United States	81 first-year social work students	Pre-post, with 3 mo postrepeated	RP counselling scenarios, central to a foundational interviewing skills course	—	Interpersonal Reactivity Index (IRI) (4 subscales : perspective taking, empathetic concern, fantasy scale, personal distress). Self-rated	IRI referenced in the literature. Reliability between subscale referenced as α = 0.61–0.81, and interitem reliabilities = 0.65–0.81.	Total empathy score overall reduced from pre (week 1) and follow-up (3 mo after completing task) ($P < 0.05$). Subscale of personal distress reduced from pre to post to follow-up ($P < 0.01$) Pre-to-post empathy decline, ES calculated as $d = 0.09$.	10.5
Grice et al,[35] 2012	United States	158 (2009 cohort), plus 126 (2010 cohort) third-year pharmacy students	Pre-post (both cohorts)	RP teaching how to use medical devices	—	Author devised rubric, subscale of Empathy Habit from Four Habits Model. Rated by faculty with live video feed of performance with SP.	None reported	Significant improvement in empathy aspects across both cohorts ($P < 0.05$) except 2009 missed one item being measured. Insufficient data to calculate ES.	12.5
Henry et al,[36] 2011	United States	127 nursing (73) and nutrition (54) students	1. RCT 2 groups with pre-post measures 2. Qualitative data	The Aging Game, (75 min including debriefing)	Class discussion (75 min)	1. Four items adapted from Maxwell and Sullivan questionnaire 2. Reflective writing Self-rated	Maxwell and Sullivan referenced in the literature (but not psychometric properties).	1. Significant improvement pre-post in intervention group in 1 of 4 items (I understand what it feels like to have problems growing older, $P < 0.05$). Intervention group significantly more empathetic than control ($P < 0.01$). ES calculated as $d = -0.24$. 2. Thematic analysis indicating cognitive and emotive aspects of empathy.	11
Hueberger et al,[37] 2010	United States	93 dietetic students	Pre-post	RP as patient with gestational diabetes (2 wk)	—	Jefferson Scale of Empathy (JSE) Self-rated	JSE adapted for dietetics students from JSPE. JSPE referenced in the literature.	Improved empathy shown in 6/20 questions (range, $P < 0.001$ to $P < 0.05$). Analysis was item by item, no ES calculated.	8.5
Kushner et al,[38] 2014	United States	127 first-year medical students	Pre-post plus 1-yr postrepeated	SP session regarding obesity	—	16-item author devised survey constructed from other instruments. Self-rated	Reliability of empathy subscale, reported as α = 0.63.	Improved empathy shown in subscale, from baseline to post ($P < .0001$) and baseline to long-term follow-up ($P = 0.001$). ES calculated as $d = 0.34$.	10.5
Lim et al,[39] 2011	New Zealand	149 fifth-year medical students, across 2 y	One year no intervention compared with next year with intervention using baseline and postmeasures	RP workshop (5 training scenarios) as part of a psychological medicine module	No intervention	Jefferson Scale of Physician Empathy Self-rated.	JSPE referenced in the literature, referenced α = 0.87–0.89.	Improved empathy ($P < 0.001$). No significant differences in baseline. ES calculated as $d = 0.57$.	13
Mawson,[47] 2014	Australia	59 second-year nursing students	Pre-post	5 min role-play with simulated auditory hallucinations	—	Author devised 11-item survey for increased knowledge, empathy, and changed attitudes toward people with auditory hallucinations Self-rated	None reported	Increase in compassion toward those who experience auditory hallucinations ($P < 0.001$). Analysis was item by item, no ES calculated.	8.5
Pacala et al,[48] 1995	United States	55 fourth-year medical students	Volunteers (39) compared with others (16), measured with pre-post	Modified Aging Game. 3-h workshop	No intervention	11 items adapted from Maxwell and Sullivan questionnaire. Self-rated.	Maxwell and Sullivan questionnaire referenced as unvalidated.	No difference in change in empathy between groups ($P = 0.15$). ES calculated as $d = 0.41$ Improved empathy by 23% ($P < 0.001$) in participants. ES calculated as $d = 0.47$.	8.5

(continued on next page)

TABLE 1. *(Continued)*

Reference	Location	Sample Size	Study Design	Simulation Modality and Topic	Comparator (Experimental Designs)	Outcome Measure/s (Empathy Only) and Rater	Psychometric Properties of Instrument	Results	MERSQI[18] (Range, 5–18)
Pacoe et al,[49] 1976	United States*	20 first-year medical students	Volunteers (13) compared with others (7), measured with pre-post	RP workshop of counselling session alternating with group discussions across 16 wk	No intervention	1. The Wells Empathic Communication Test (WECT) 2. The Index of Facilitative Discrimination (IFD) (recognition of empathy) Raters assessed written responses to 10 videos.	WECT and IFD psychometric testing referenced in the literature.	Improved scores for intervention group with both measures ($P < 0.001$ for both) ES calculated as $\Theta = 2.96$ (nonparametric)	11.5
Sanson-Fisher et al,[40] 1980	Australia*	40 second-year medical students	RCT, 2-group cross-over study, single measure of response to intervention	SP, history taking with patients with neurotic disorders	Real patients	Truax empathy scale Trained assessors of random 3 × 2-min selection of audio recording, average ratings used.	Truax empathy scale referenced in the literature Concordance between 20% randomly selected segments and original rating, 93.8%	No significant improvement in empathy between simulated and real patients in total or in either arm of crossover. ES real-simulated calculated as $d = 0.25$ (favors real) ES simulated-real calculated as :$d = 0.25$ (favors real)	
Schweller et al,[1] 2014	Brazil	124 fourth-year medical students n = 123 fifth-year medical students)	Pre-post (fourth-year students administered posttest after 3 mo); sixth-year students administered posttest after 1 mo	Each learner experiences an SP encounter, plus observation of others and debriefing; emphasis on patient feelings	—	Jefferson Scale of Physician Empathy (JSPE), IRI. Self-rated	JSPE and IRI referenced in the literature.	Fourth-year students demonstrated a significant increase in JSPE ($P < 0.001$, ES = 0.61) and IRI ($P = .003$, ES = 0.19). Sixth-year students demonstrated a significant increase in JSPE ($P < 0.001$, ES = 0.64) and IRI ($P < 0.001$, ES = −0.20).	9
Vannatta et al,[41] 1996	United States	154 first-year medical students	RCT, 3 groups with pre-post and postmeasures	SPs with feedback from SPs teaching medical interviewing	SPs with feedback from faculty and RP with feedback from faculty	Empathy skills from (1) modified RIAS—pre-post measures and (2) Arizona Clinical Interviewing Rating Scale (ACIRS)—post only. Trained assessors rated video-recorded SP encounters.	RIAS and ACIR referenced in the literature.	SPs with feedback from SPs—ACIR significantly improved relative to feedback to faculty alone ($P = 0.2$, calculated d = 0.60). RIAS no significant difference (calculated $d = 0.024$ (favors faculty feedback). SPs with SP feedback and RP with faculty feedback. ACIR, no significant difference [calculated $d = 0.0822$ (favors SP feedback)] RIAS no significant difference [calculated $d = 0.3791$ (favors SP feedback)]. SP with faculty feedback and RP with faculty feedback. ACIR, no significant difference [calculated $d = 0.5153$ (favors RP)]; RIAS, no significant difference $d = 0.403$ (calculated d favors faculty feedback).	12.5
Varkey et al,[50] 2006	United States*	84 first-year medical students	Pre-post	Modified Aging Game	—	11 items from the modified Maxwell and Sullivan questionnaire. Self-rated	Maxwell and Sullivan questionnaire referenced as unvalidated.	Improved attitudinal change in 6/8 questions (range, $P = 0.0001$ to $P = 0.049$); increased empathy in 2 of 3 questions ($P < 0.0001$ and $P < 0.0002$) Analysis was item by item, no ES calculated.	9.5

*Location, if not specified, is the location of the primary author's affiliation.
Calculated, statistic compiled by review authors.
d, Cohen *d;* ES, effect size; RP, role-play methodology; SP, simulated patient methodology; α, Cronbach α.

TABLE 2. Details of Randomized Controlled Trials

Study Details	RCT Design	Comparisons	Outcome Measure/s (Empathy Only)	Rater	Significance	Learner Role in Simulation (BP or BHP)	Effect Size favors Simulation, Unless Noted Otherwise
Bosse et al,[25] 2012 Germany 103 fifth-year medical students SP or RP communication with and counselling of parents of sick children (9 cases)	Three groups using pre-post measures	1. RP with alternative course 2. SP with alternative course 3. RP with SP	Calgary-Cambridge subscale—previously validated communication skills scale External behavioral rating.	Observational rating within OSCE	1. RP significantly improves empathy compared with alternative. 2. SP significantly improves empathy compared with alternative. 3. RP significantly improves empathy compared with SP.	1. BP and BHP 2. BHP only 3. Comparing 1 and 2	1. $d = 1.47$ 2. $d = 0.46$ 3. $d = 1.12$ (favors RP)
Bunn et al,[26] 2009 United States 150 medical students 40-min simulated auditory hallucination	Two-group RCT using pre-post measures	Simulation with no intervention	JSPE-S—previously validated empathy specific scale	Self-rated	Simulation significantly improves empathy	BP	$d = 0.60$
Cahan et al,[21] 2010 (study 1) SP following communication skills training, workshop part of a human factors curriculum	Two-group RCT using postmeasures	Simulation with no intervention	Author devised assessment rubric grading written response to vignettes. α reported as 0.808	Rating of written responses	No significant improvement in empathy compared with control	BHP	$d = 0.19$
Daeppen et al,[29] 2012 Switzerland Fifth-year medical students Motivational interviewing after training 2 × 4-h workshops, with significant RP component	Two-group RCT using postmeasures	Motivational interviewing workshops with no motivational interviewing workshops	Motivational Interviewing Treatment Integrity (MITI) scoring tool empathy subscale – previously validated.	Videos of SPs by trained raters	Empathy subscale significantly higher in intervention group	BP/BHP	$d = 0.92$
Deladisma et al,[30] 2007 United States 84 medical students Virtual patient case with abdominal pain	Two-group randomized trial, single measure of response to intervention	Virtual patient with simulated patient	Consensus devised score sheet including observed behavior and specific empathy item. α r reported as 0.92.	Videos rated by clinicians for nonverbal communication	SP significant improves empathy relative to virtual patient.	BHP	$d = 0.70$ (favors SP)
Gleber et al,[3] 1995 United States 24 dental hygiene students SP as part of other modalities within 20 h on interpersonal skills training	Two-group RCT with pre-post and 12-mo follow-up measures	Interpersonal skills training with no interpersonal skills training	Mehrabian and Epstein Measure of Emotional Empathy (previously validated)	Self-rated	No significant improvement in empathy in intervention compared with control group or at follow-up.	BHP	$d = 0.11$
Henry et al,[36] 2011 United States 127, nursing (73) and nutrition (54) students The Aging Game, (75 min including debriefing)	Two-group RCT with pre-post measures	Simulation with class discussion (75 min)	4 items adapted from Maxwell and Sullivan (referenced but not validated).	Self-rated	Intervention group significantly more empathetic than control.	BP	$d = 0.24$

(continued on next page)

TABLE 2. *(Continued)*

Study Details	RCT Design	Comparisons	Outcome Measure/s (Empathy Only)	Rater	Significance	Learner Role in Simulation (BP or BHP)	Effect Size favors Simulation, Unless Noted Otherwise
Sanson-Fisher et al,[40] 1980 Australia 40 second-year medical students History taking with SP/patients with neurotic disorders	Two-group RCT crossover study, single measure of response to intervention	1. Real patients with SPs 2. SPs with real patients	Truax empathy scale (previously validated).	Trained raters on 6 min of audio, from a combination of assessment on both real patients and SPs	No significant improvement in empathy between simulated and real patients in total or in either arm of crossover.	BHP	$d = 0.30$ (favors real patient)
Vannatta et al,[41] 1996 United States 154 first-year medical students SPs with feedback from SPs teaching medical interviewing	Three-group RCT with pre-post and postmeasures	1. SPs with feedback from SPs and SPs with feedback from faculty 2. SPs with feedback from SPs and RP with feedback from faculty 3. SPs with feedback from faculty and RP with feedback from faculty	Empathy skills from: 1. modified RIAS—psychometric 2. Arizona Clinical Interviewing Rating Scale (ACIRS), both previously validated in literature.	Videos of SPs assessed by trained raters	SPs with feedback from SPs—ACIR significantly improved relative to feedback to faculty alone; no other significant differences between categories.	1. BHP only 2. BHP compared with BHP and BP 3. BHP compared with BHP and BP	1. ACIR $d = 0.60$ (favors SP/SP feedback), RIAS $d = 0.02$ (favors SP/faculty feedback) 2. ACIR $d = 0.08$ (favors SP/SP feedback), RIAS $d = 0.34$ (favors SP/SP feedback) 3. ACIR $d = 0.52$ (favors RP/faculty feedback), RIAS $d = 0.40$ (favors SP/faculty feedback)

BHP, Being a health professional; BP, Being a patient; *d*, Cohen *d*; RP, role-play.

student empathy development between simulation intervention and a control. In all of these, the intervention required the learners to only "be a health professional" and never a patient. All 4 RCTs that contained interventions where the learner assumed the role of the patient all or some of the time within the simulation[22,24,27,28] showed significant improvements relative to a control. These are represented in the forest plot in Figure 2. The study of Sanson-Fisher et al,[23] that compared SP and real patient encounters, is separated in the forest plot, as the comparator is very different. RCTs that compare different aspects of simulation are not included in this figure.

There was a range of educational design features surrounding the simulated experience. There were 4 (15%) of 27 comparative studies that compared different types and features of simulation in promoting empathy; 3 of these were randomized (Table 2). Two of these indicated that role-play encounters improved learner's empathy more than SP encounters[27] or using a mannequin.[43] One study indicated that SP encounters improved learner's empathy more than virtual patients,[31] and another study had mixed results regarding the efficacy of feedback from SPs compared with feedback from faculty.[46]

DISCUSSION

The findings of this review suggest that simulation may be an appropriate educational methodology for developing empathy and/or empathetic behaviors in preservice health professional students. This finding was not universal, and it seems dependent on the type and educational features of the simulations as well as the definition of empathy and associated measures. The variety and complexity of the reported educational designs are exciting from a teaching perspective but make definite conclusions challenging from a research perspective. The challenges with measures are brought into focus by 1 study,[29] which indicated a decline in empathy after the intervention. The authors attributed this to a decline in the "personal distress" subscale of the Interpersonal Reactivity Index, which measures "one's own feelings of discomfort in the face of emotionally challenging interpersonal situations."[29] This inconsistency highlights broader debates about whether empathy should be measured and taught as an internal process or as observable communications.[32]

Collectively, the randomized controlled studies suggest that the simulation approach that seems most beneficial is one that asks the learners to literally stand in patients' shoes.

These results are also supported by an RCT, which specifically examined the issue of "playing the patient." In this study, Bosse et al[27] compared communication skills development in learners role-playing (that is, acting as both learner and health professional) with those working with a simulated patient (that is, acting as a health professional) and with a control (no simulation experience). Final-year medical students (n = 103) were assigned to 3 groups receiving education and training in counseling caregivers of sick children. Two groups received counseling and communication training using either role-plays or simulated patient encounters, whereas the control group received the only standard coursework learning material. After the

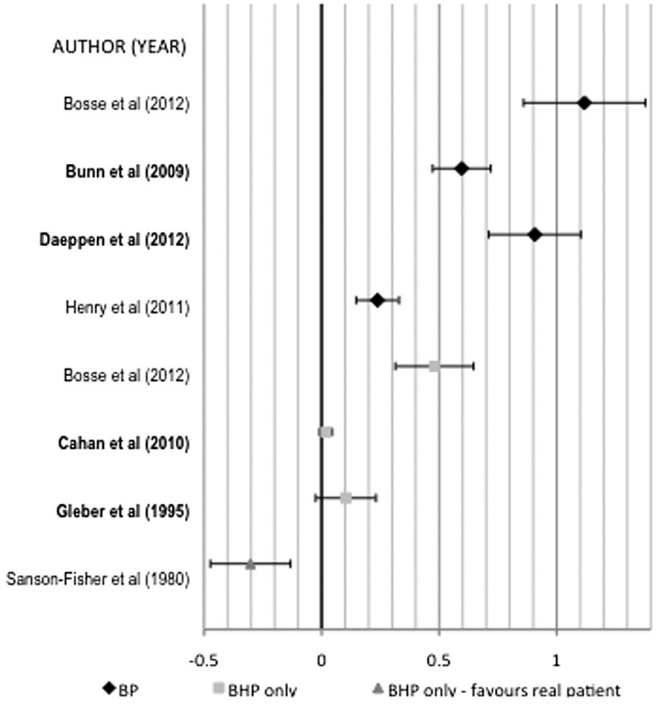

FIGURE 2. Forest plot of RCTs that compare simulation to no or some alternative, with empathy as an outcome. Effect sizes with 95% confidence interval are shown in 3 themes, represented by diamonds, squares, and a triangle. Effect size of RCTs where the comparison is between a ''being the patient'' (including role-plays) simulation and some or no alternative. Effect size of RCTs where the comparison is between a ''being health professional'' only (excluding role plays) and some or no alternative (excluding real patients). Effect size of RCT where the comparison is between a ''being a health professional'' simulation and real patients. Across these themes, the study author and date are in bold when the comparator is ''no alternative''; other studies have nonsimulation comparators.

interventions, 6 OSCE stations were undertaken, each with 1 SP. Final OSCE scores were rated using the Calgary-Cambridge Referenced Observation Guide (CCROG); analysis indicated that the role-play group significantly improved compared with the SP group with respect to "understanding the patient's perspective" ($t = 5.11$, and $P < 0.001$, $d = 1.12$), and notably, this was the only domain where this significant difference was found. This study was rated 15.5/18 on the MERSQI.

The theoretical constructs of empathy support this idea. Assuming the role of the patient introduces the understanding and shared feelings of the patient's perspective, but then, the learner must remove themselves from the simulation and through debriefing or feedback processes and must translate this experience into the empathetic behaviors we expect from health care practitioners. Simulation education might combat the decline of empathy noted in later years of study,[16] as it permits students to manage both a less idealistic view of health care practice and an appropriate level of identification with patients.

Role-play may be particularly valuable. Role-play permits rotating roles of patient and health care provider and therefore provides both the experience of the patient and the experience of working with patients. This role reversal may be an important mechanism in developing empathy and deserves further exploration. Role-play is sometimes seen as a less desirable but a cheaper alternative for SP encounters, but perhaps, it could become first choice for learning empathy and empathetic behaviors. There is also a caveat to using role-play methodology. Without creating an appropriately safe learning environment,[33] acting as a surrogate for a patient might also provoke other emotions such as vulnerability or anxiety.[35] The role of debriefing may be particularly important in assisting students to translate their experiences as role-play patients to general communication skills.

It is also worth noting that of the 17 studies where the learners were "being a health professional," only 1 focussed solely on empathy development. Empathy development was interwoven with learning other skills such as interviewing,[31] motivational interviewing,[24,41] effective patient/client communication skills,[3,21,27,29] interpersonal skills,[3] and psychomotor skills.[43] In many instances, these associated skills developed alongside empathy; sometimes, these skills were used as markers or indicator of empathy. For example, Deladisma et al[31] reported the development of nonverbal communication skills and learning to ask clear questions as a measure for increased empathy.

This review builds on and adds to the work of other related systematic reviews, particularly those of Stepien and Baernstein[37] and Batt-Rawden et al.[13] It contains 23 additional studies not reported within these previous reviews, which explore empathy in medical education, and specifically draws the link between empathy and simulation. This review highlights the value of taking the role of patient, either through role-play or as part of specific "patient

experience" simulation design, which was not noted in previous studies, although it is congruent with their findings.

The limitations to the findings of this review are derived from the included studies and from the review process itself. With respect to the included studies, although some had high-quality experimental designs, these were in the minority. Reporting standards were variable; effect sizes were rarely reported, and sometimes, even a lack of means and SDs made calculating these statistics challenging. There were inconsistencies with the outcome measures, as noted earlier in the discussion. There was notably a lack of experimental data examining groups over periods, which might give information about how empathy decays after interventions. The nature of the comparator was not well considered. As has been noted previously,[38] studies that aim to compare a new medium to some or no alternatives have limitations because of the difficulty in forming valid comparison groups; this type of the design formed sizeable majority of the included studies. More consideration could be given as to when and why simulation is the optimal way to teach empathy, building on the body of work reported in this review.

Future work might include further experimental studies to replicate the results of the study of Bosse et al,[27] which compared the effects of empathy on SP encounters relative to role-play encounters and would be particularly useful, as well as in-depth qualitative comparative investigations of the learner experience "being a patient." It would be interesting to see if this type of empathy education might be effective for practicing health professionals. The role of feedback is worthy of further exploration. In addition, with the exception of 1 study comparing student behavior with simulated to real patients, all other behavioral ratings were not assessed with real patient encounters, with most tested in simulation. Practitioners have had improvements in empathy assessed by real patients,[14,15] and we suggest this type of assessment could be extended to preservice health practitioners.

The review process itself also had limitations. The synthesis was based on a simple categorization of studies. The review strategy was less likely to include qualitative studies, as few of these are framed as comparative studies. There was no systematic hand-searching for additional articles. Balanced against this is the comprehensive search strategy that incorporated 11 databases, independent review for inclusion of each full text by 2 authors, use of an identical data extraction table for all studies, independent duplicate quality assessment, and a synthesis approach that provides an insight into possible learning mechanisms.

In conclusion, the findings of this review suggest that simulation may be a useful educational methodology for developing empathetic behaviors in preservice health professional students. The most interesting inference is the notion that simulations that ask the learner to act in the role of patient may be more effective in developing empathy. This may have significant implications for educational design of simulations in preservice health professional curricula. Further research is needed to confirm this result and also to investigate other features of simulation, which promote or inhibit learning empathy.

ACKNOWLEDGMENTS

The authors thank Anne Young, expert librarian, for her guidance and assistance in searching the literature, and the anonymous reviewers of this article.

REFERENCES

1. Schweller M, Costa FO, Antônio MARGM, Amaral EM, De Carvalho-Filho MA. The impact of simulated medical consultations on the empathy levels of students at one medical school. *Acad Med* 2014;89(4):632–637.
2. Dearing KS, Steadman S. Challenging stereotyping and bias: a voice simulation study. *J Nurs Educ* 2008;47(2):59–65.
3. Gleber JM. Interpersonal communications skills for dental hygiene students: a pilot training program. *J Dent Hyg* 1995;69(1):19–30.
4. Ker J, Bradley P. Simulation in m edical e ducation. In: Swanick T, ed. *Understanding Medical Education: Evidence, Theory and Practice*. Chicester, England: Wiley-Blackwell, 2010:164–180.
5. Wear D, Varley JD. Rituals of verification: the role of simulation in developing and evaluating empathic communication. *Patient Educ Couns* 2008;71(2):153–156.
6. de la Croix A, Skelton J. The simulation game: an analysis of interactions between students and simulated patients. *Med Educ* 2013;47(1):49–58.
7. Hojat M. Definitions and conceptualization. In: *Empathy in Patient Care: Antecedents, Development, Measurement, and Outcomes*. New York, NY: Springer; 2007:3–15.
8. Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. *J Pers Soc Psychol* 1983;44(1):113.
9. Mercer SW, McConnachie A, Maxwell M, Heaney D, Watt GC. Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Fam Pract* 2005;22(3):328–334.
10. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med* 2011;86(3):359–364.
11. Del Canale S, Louis DZ, Maio V, et al. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. *Acad Med* 2012;87(9):1243–1249.
12. Yang C-PP, Hargreaves WA, Bostrom A. Association of empathy of nursing staff with reduction of seclusion and restraint in psychiatric inpatient care. *Psychiatr Serv* 2014;65(2):251–254.
13. Batt-Rawden SA, Chisolm MS, Anton B, Flickinger TE. Teaching empathy to medical students: an updated, systematic review. *Acad Med* 2013;88(8):1171–1177.
14. Riess H, Kelley J, Bailey R, Dunn E, Phillips M. Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. *J Gen Intern Med* 2012;27(10):1280–1286.
15. Bonvicini KA, Perlin MJ, Bylund CL, Carroll G, Rouse RA, Goldstein MG. Impact of communication training on physician expression of empathy in patient encounters. *Patient Educ Couns* 2009;75(1):3–10.
16. Neumann M, Edelhauser F, Tauschel D, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Acad Med* 2011;86(8):996–1009.
17. Nunes P, Williams S, Sa B, Stevenson K. A study of empathy decline in students from five health disciplines during their first year of training. *Int J Med Educ* 2011;2: 12–17.
18. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *J Gen Intern Med* 2008;23(7):903–907.
19. Wong G, Greenhalgh T, Pawson R. Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Med Educ* 2010;10(1):12.

20. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to not get your paper published. *Adv Health Sci Educ Theory Pract* 2014;19(1):1–5.

21. Cahan MA, Larkin AC, Starr S, et al. A human factors curriculum for surgical clerkship students. *Arch Surg* 2010;145(12):1151–1157.

22. Bunn W, Terpstra J. Cultivating empathy for the mentally ill using simulated auditory hallucinations. *Acad Psychiatry* 2009;33(6):457–460.

23. Sanson-Fisher RW, Poole AD. Simulated patients and the assessment of medical students' interpersonal skills. *Med Educ* 1980;14(4):249–253.

24. Daeppen JB, Fortini C, Bertholet N, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. *Patient Educ Couns* 2012;87(3):313–318.

25. Pacala JT, Boult C, Bland C, O'Brien J. Aging game improves medical students' attitudes toward caring for elders. *Gerontol Geriatr Educ* 1995;15(4):45–57.

26. Galletly CA, Burton CA. The effects of experiencing simulated auditory hallucinations on attitudes to schizophrenia in final year medical students. *Schizophr Res* 2010;117(2–403):402–3.

27. Bosse HM, Schultz JH, Nickel M, et al. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. *Patient Educ Couns* 2012;87(3):300–306.

28. Henry BW, Ozier AD, Johnson A. Empathetic responses and attitudes about older adults: how experience with the aging game measures up. *Educ Gerontol* 2011;37(10):924–941.

29. Gockel A, Burton DL. An evaluation of prepracticum helping skills training for graduate social work students. *J Soc Work Educ* 2014;50(1):101–119.

30. Varkey P, Chutka DS, Lesnick TG. The aging game: improving medical students' attitudes toward caring for the elderly. *J Am Med Dir Assoc* 2006;7(4):224–229.

31. Deladisma AM, Cohen M, Stevens A, et al. Do medical students respond empathetically to a virtual patient? *Am J Surg* 2007;193(6):756–760.

32. Greason PB, Cashwell CS. Mindfulness and counseling self-efficacy: the mediating role of attention and empathy. *Counselor Education and Supervision* 2009;49(1):2–19.

33. Baile WF, Blatner A. Teaching communication skills: using action methods to enhance role-play in problem-based learning. *Simul Healthc* 2014;9(4):220–227.

34. Bath LE, Cunningham S, McIntosh N. Medical students' attitudes to caring for a young infant—can parenting a doll influence these beliefs? *Arch Dis Child* 2000;83(6):521–523.

35. Nestel D, Tierney T. Role-play for medical students learning about communication: guidelines for maximising benefits. *BMC Med Educ* 2007;7: 3.

36. Fields SK, Mahan P, Tillman P, Harris J, Maxwell K, Hojat M. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider—student version. *J Interprof Care* 2011;25(4):287–293.

37. Stepien KA, Baernstein A. Educating for empathy. A review. *J Gen Intern Med* 2006;21(5):524–530.

38. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med* 2005;80(6):541–548.

39. Kushner RF, Zeiss DM, Feinglass JM, Yelen M. An obesity educational intervention for medical students addressing weight bias and communication skills using standardized patients. *BMC Med Educ* 2014;14: 53.

40. Heuberger R. Increasing empathy, knowledge, and skills regarding gestational diabetes. *Topics in Clinical Nutrition* 2010;25(3): 250–255.

41. Lim BT, Moriarty H, Huthwaite M. "Being-in-role": a teaching innovation to enhance empathic communication skills in medical students. *Med Teach* 2011;33(12):e663–e669.

42. Chaffin AJ, Adams C. Creating empathy through use of a hearing voices simulation. *Arch Dis Child* 2013;9(8):e293–e304.

43. Chunharas A, Hetrakul P, Boonyobol R, Udomkitti T, Tassanapitikul T, Wattanasirichaigoon D. Medical students themselves as surrogate patients increased satisfaction, confidence, and performance in practicing injection skill. *Med Teach* 2013;35(4):308–313.

44. Evans S, Lombardo M, Belgeri M, Fontane P. The geriatric medication game in pharmacy education. *Am J Pharm Educ* 2005;69(3):304–310.

45. Mawson K. Use of media technology to enhance the learning of student nurses in regards to auditory hallucinations. *Int J Ment Health Nurs* 2014;23(2):135–144.

46. Vannatta JB, et al. Comparison of standardized patients and faculty in teaching medical interviewing. *Acad Med* 1996;71(12):1360–1362.

