

# Computerized Virtual Patients in Health Professions Education: A Systematic Review and Meta-Analysis

David A. Cook, MD, MHPE, Patricia J. Erwin, MLS, and Marc M. Triola, MD

## Abstract

### Purpose

Educators increasingly use virtual patients (computerized clinical case simulations) in health professions training. The authors summarize the effect of virtual patients compared with no intervention and alternate instructional methods, and elucidate features of effective virtual patient design.

### Method

The authors searched MEDLINE, EMBASE, CINAHL, ERIC, PsychINFO, and Scopus through February 2009 for studies describing virtual patients for practicing and student physicians, nurses, and other health professionals. Reviewers, working in duplicate, abstracted information on instructional

design and outcomes. Effect sizes were pooled using a random-effects model.

### Results

Four qualitative, 18 no-intervention controlled, 21 noncomputer instruction-comparative, and 11 computer-assisted instruction-comparative studies were found. Heterogeneity was large ( $I^2 > 50\%$ ) in most analyses. Compared with no intervention, the pooled effect size (95% confidence interval; number of studies) was 0.94 (0.69 to 1.19; N=11) for knowledge outcomes, 0.80 (0.52 to 1.08; N=5) for clinical reasoning, and 0.90 (0.61 to 1.19; N=9) for other skills. Compared with noncomputer instruction, pooled effect size (positive numbers favoring virtual patients) was  $-0.17$  ( $-0.57$  to  $0.24$ ;

N=8) for satisfaction,  $0.06$  ( $-0.14$  to  $0.25$ ; N=5) for knowledge,  $-0.004$  ( $-0.30$  to  $0.29$ ; N=10) for reasoning, and  $0.10$  ( $-0.21$  to  $0.42$ ; N=11) for other skills. Comparisons of different virtual patient designs suggest that repetition until demonstration of mastery, advance organizers, enhanced feedback, and explicitly contrasting cases can improve learning outcomes.

### Conclusions

Virtual patients are associated with large positive effects compared with no intervention. Effects in comparison with noncomputer instruction are on average small. Further research clarifying how to effectively implement virtual patients is needed.

**D**iagnostic errors represent a significant source of patient morbidity, and cognitive errors represent the most common cause of diagnostic error.<sup>1-4</sup> Cognitive errors also lead to suboptimal treatment decisions.<sup>3</sup> Evidence suggests that most cognitive errors arise from faulty interpretation, synthesis, and judgment rather than insufficient data gathering or fund of knowledge<sup>2,3,5</sup> and that decreasing the incidence of cognitive error will require that health care

providers experience multiple, varied patient cases.<sup>6,7</sup> Yet even as the rapid growth of medical information and expectations for quality care have increased the complexity of medical decision making, we see decreased time for education<sup>8</sup> and heightened concerns regarding patients as educational subjects.<sup>9</sup> Safer and more efficient means of facilitating the development of clinically relevant knowledge and skills are needed. The computer-screen-based virtual patient, "a specific type of computer program that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions,"<sup>10</sup> has been proposed as one way to develop these essential cognitive clinical skills.<sup>4,11-13</sup>

Educators would benefit from a better understanding of the potential effectiveness of virtual patients in health professions training, the design features commonly employed when implementing virtual patients, and which of these features are associated with

improved learning outcomes.<sup>14</sup> A review and synthesis of evidence from existing studies could inform decisions on when and how to effectively use virtual patients. Although previous reviews have focused on surgical and procedural simulators,<sup>15-18</sup> we are aware of only one review of research on virtual patients.<sup>19</sup> This review had important methodological limitations, including incomplete accounting of existing studies, limited assessment of study quality, and no quantitative pooling of study results. Other reviews of computer-assisted instruction have incorporated research on virtual patients as only a small minority of included studies.<sup>20-26</sup> In the present review, we sought to identify and summarize all studies involving virtual patients for training health professionals.

**Dr. Cook** is professor of medicine and director, Office of Education Research, College of Medicine, Mayo Clinic, Rochester, Minnesota.

**Ms. Erwin** is assistant professor of medical education, College of Medicine, Mayo Clinic, Rochester, Minnesota.

**Dr. Triola** is assistant professor of medicine and chair, Division of Educational Informatics, New York University School of Medicine, New York, New York.

Correspondence should be addressed to Dr. Cook, Division of General Internal Medicine, Baldwin 4-A, 200 First Street SW, Rochester, MN 55905; telephone: (507) 266-4156; e-mail: cook.david33@mayo.edu.

*Acad Med.* 2010;85:1589-1602.  
First published online August 10, 2010  
doi: 10.1097/ACM.0b013e3181edfe13

Supplemental digital content is available for this article at <http://links.lww.com/ACADMED/A23>. Clickable links to the material are provided in the HTML text and PDF of this article at [www.academicmedicine.org](http://www.academicmedicine.org).

## Method

This review was planned, conducted, and reported in adherence to standards of quality for reporting meta-analyses (QUOROM, MOOSE, and PRISMA).<sup>27–29</sup> No ethical approval was needed, as the study did not involve human participants.

## Question

We sought to answer two questions: How effective are virtual patients in comparison with no intervention and alternate instructional methods, and what virtual patient design features are associated with higher learning outcomes?

## Data sources and searches

We included studies published in any language that investigated use of a virtual patient to teach health professions learners (students, postgraduate trainees, or practitioners in a profession directly related to human or animal health, including physicians, dentists, nurses, pharmacists, veterinarians, and physical therapists) at any stage in training or practice. We defined a virtual patient as “a specific type of computer program that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions.”<sup>10</sup> This excluded other forms of computer-based learning in which patient cases did not require the user to interactively gather patient data, and other forms of simulation such as standardized patients, manikins, part-task trainers, and systems requiring specialized equipment not found on a typical personal computer. We excluded studies using computer cases for which replication would require hardware not included with a typical personal computer (e.g., haptics input devices or virtual reality head-mounted displays) because such cases differ from virtual patients in educational objectives and instructional methods. We also excluded computer simulations used for procedural planning or disease modeling, simulations that did not involve obtaining a history or exam, and computer-mediated consultations on real patients. Some early studies (i.e., before the ability to display computer graphics) used slide projectors and videotape recorders as part of the virtual environment. We included such studies

provided that the external visual system was fully integrated with a computerized virtual patient, and a modern personal computer would be able to replicate the system without special equipment. We made no exclusions based on outcome type.

An experienced research librarian (P.J.E.) designed a strategy to search MEDLINE, EMBASE, CINAHL, ERIC, PsychINFO, Scopus, and the University of Toronto Research and Development Resource Base using search terms including “virtual patient,” “computer simulation,” “problem-based learning,” “case-based learning,” “clinical simulation,” and “medical education” (see Supplemental Digital List 1, available at <http://links.lww.com/ACADMED/A23>). We used no beginning date cutoff, and the last date of search was February 16, 2009. We identified additional studies by searching reference lists of all included articles.

## Study selection

We worked independently and in duplicate to screen all titles and abstracts for inclusion. In the event of disagreement or insufficient information in the abstract, we reviewed the full text of potential articles, again independently and in duplicate. We resolved conflicts by consensus. Chance-adjusted interrater agreement for study inclusion, determined using intraclass correlation coefficient<sup>30</sup> (ICC), was 0.69.

## Data extraction and quality assessment

We developed a data abstraction form through iterative testing and revision. We abstracted data independently and in duplicate for all variables where reviewer judgment was required, using ICC to determine interrater agreement and resolving conflicts by consensus.

We classified studies as descriptive (description of a virtual patient with only a single-group postintervention evaluation or no evaluation), no-intervention controlled (single-group pretest–posttest comparison, or comparison with another group receiving no intervention), media-comparative (comparison with a group receiving a noncomputer educational intervention), or computer-assisted learning comparative (comparison with a group receiving a computer-assisted educational intervention, including an alternate

virtual patient) (ICC = 0.83). We also identified studies with rigorous qualitative analysis (ICC = 0.74). For all studies, we abstracted information on

- the training level of learners,
- the clinical topic,
- the method of interview (learners interrogated the virtual patient using free text [*natural language*] or using a predefined *menu* of questions; ICC = 0.84),
- the type of case progression (*free navigation* [patient status remained essentially unchanged as learner gathers information], *linear* [patient evolved over time, but followed the same course regardless of learner decisions], or *branching* [patient evolved with learner decisions affecting subsequent events]; ICC = 0.65),
- the presence of learner collaboration (learners completed cases *alone* or as a *group*; ICC = 0.70), and
- the type(s) of outcome evaluated<sup>20</sup>: satisfaction, knowledge/attitudes, clinical reasoning, skills, and/or behaviors in practice or effects on patient care (ICC range 0.74–0.96).

For quantitative comparative studies, we coded, for both the virtual patient and the comparison, the presence of features of effective simulations identified in a review of simulation<sup>15</sup> (ICC range 0.58–0.83), namely:

- feedback provided to learner (*low*, *moderate*, or *high*),
- opportunity for repetitive practice (*present/absent*),
- curriculum integration (virtual patient was an integrated part of the curriculum/course [*present*], or an optional activity [*absent*]),
- range of task difficulty (some cases designed to be harder than others; *present/absent*),
- multiple learning strategies (*few* [0–1 strategies], *moderate* [2–3]), or *many* [ $\geq 4$ ] strategies, such as providing worked examples, offering hints, enabling group discussion, giving feedback, or requiring learners to list a full differential diagnosis, justify choices, or explicitly contrast different cases),

- clinical variation (cases reflected a spectrum of illness states; *present/absent*), and
- individualized learning (how well the system adapted to individual learning needs: *low, moderate, or high*; branching navigation was considered moderate).

We also coded the amount of interactivity (degree to which the course design encouraged learners to cognitively engage<sup>20</sup>; ICC = 0.71), the time spent learning (ICC = 0.85), and the number of cases or practice problems. We abstracted information on study design (number of groups [ICC = 0.99], method of group assignment [ICC = 0.72], and timing of assessments [pretest–posttest versus posttest-only; ICC = 0.87]), outcomes (subjective/objective [ICC = 0.63–1.0]), quantitative outcome results, and methodological quality. Methodological quality was graded using an adaptation of the Newcastle–Ottawa Scale for cohort studies<sup>20,31</sup>: representativeness of the intervention group (ICC = 0.87), selection of the comparison group (ICC = 0.37), comparability of cohorts (statistical adjustment for baseline characteristics in nonrandomized studies [ICC = 0.58], or randomization [ICC = 0.88] and allocation concealment for randomized studies [ICC = 0.27]), blinding of outcome assessment (ICC = 0.47–0.91), and completeness of follow-up (ICC = 0.30–0.73).

For qualitative studies, we abstracted the information sources, analytic method, and main themes identified.

### Data synthesis

We abstracted information separately for outcomes of satisfaction, knowledge, clinical reasoning, skills, and behaviors/patient effects. We converted each mean and standard deviation (SD) or odds ratio to a standardized mean difference (Hedges *g* effect size [ES]).<sup>32–34</sup> When this information was unavailable, we estimated the ES using statistical test results (e.g., *P* values).<sup>32</sup> For two-group pretest–posttest studies, we used posttest means adjusted for pretest or adjusted statistical test results, or, if these were not available, we standardized the difference in change scores using the pretest variance.<sup>33</sup> For crossover studies, we used means or exact statistical test results adjusted for repeated measures, or, if

these were not available, we used means pooled across each intervention.<sup>35,36</sup> For one study reporting neither *P* values nor any measure of variance, we used the average SD from all other included studies.

We used the  $I^2$  statistic<sup>37</sup> to quantify inconsistency (heterogeneity) across studies.  $I^2$  estimates the percentage of variability across studies not due to chance, and values >50% indicate large inconsistency. Because we found large inconsistency in most analyses, we used random-effects models to pool weighted ESs using StatsDirect 2.6.6 ([www.statsdirect.com](http://www.statsdirect.com)). Sensitivity analyses using fixed-effects models and excluding the study using imputed SDs provided virtually identical results, and we do not report them here.

Acknowledging that technologies evolve and earlier studies may no longer apply to modern computer interventions, we conducted sensitivity analyses excluding studies prior to 1991 (the year in which the World Wide Web was first described). We conducted subgroup analyses based on study design (one-group pretest–posttest versus two-group for no-intervention-controlled studies, and randomized versus nonrandomized for active-intervention-controlled studies), outcome blinding, total quality score (above versus at/below median), and instructional design (amount of interactivity, amount of feedback, number of learning strategies, and time spent learning). To increase power for subgroup analyses, we collapsed knowledge, clinical reasoning, and skill outcomes to a single outcome of “performance,” with the latter outcomes taking precedence in cases of multiple outcomes per study. Although funnel plots can be misleading in the presence of inconsistency,<sup>38</sup> we used these along with the Egger asymmetry test<sup>39</sup> to explore evidence of publication bias.\* In cases of

\*Funnel plots graph each study’s effect size against the study’s sample size in attempt to discern whether small studies have been left unpublished because they failed to show statistically significant results (publication bias). Asymmetric funnel plots suggest (but do not confirm) publication bias, while symmetric plots suggest (but do not guarantee) its absence. The trim and fill method attempts to balance an asymmetric plot in order to determine a more trustworthy (unbiased) effect size estimate. However, both the funnel plot and the trim and fill method have important limitations, as noted in the references cited above. Results of both methods should be considered at best tentative or suggestive.

asymmetry, we used the trim and fill method to estimate revised pooled ES estimates, although this method also has limitations when inconsistency is present.<sup>40</sup> We used SAS 9.1 (SAS Institute, Cary, North Carolina) to calculate ICC. Statistical significance was defined by a two-sided alpha of .05, and interpretations of clinical significance emphasized confidence intervals (CIs)<sup>41</sup> in relation to Cohen ES classifications.<sup>42</sup>

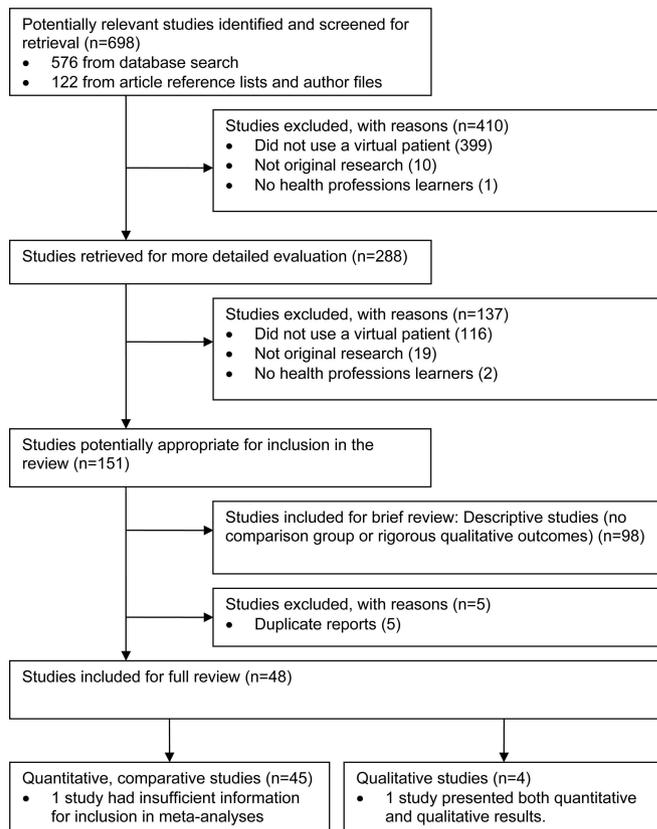
We synthesized qualitative studies by identifying key themes and supportive statements, initially independently in duplicate and then by consensus, and iteratively revising and reclassifying these themes.

## Results

### Trial flow

We identified 576 citations using our search strategy and an additional 122 articles from author files and reviews of reference lists. From these, we identified 151 potentially eligible articles (Figure 1), of which 98 reported descriptions without comparison or qualitative analysis. Of the other 53, 5 were duplicate reports of a previously reported study (see Supplemental Digital List 2 at <http://links.lww.com/ACADMED/A23> for listing), and we selected the most detailed report for full review. Forty-five articles<sup>43–87</sup> reported comparisons with no intervention (N = 18), another instructional format (N = 21), or another virtual patient (N = 11) (some articles reported more than one comparison). We successfully contacted authors of 5 articles for additional outcomes information and received information from 4. Still, one otherwise eligible article contained insufficient data to calculate an ES and was excluded from the meta-analyses. However, this article, along with 3 others, reported rigorous qualitative analyses.<sup>60,88–90</sup>

The earliest study we identified, published in 1966,<sup>43</sup> used the University of Illinois PLATO system to implement a virtual patient for nurses. Reports describing a system for surgeons at University of Leeds<sup>91</sup> and the TIME system<sup>92</sup> for medical students followed shortly thereafter. Since then, virtual patients have been used to teach topics such as chest pain, endodontics, motivational interviewing, psychiatric



**Figure 1** Trial flow for a systematic review of computerized virtual patients. Supplemental material contains details on descriptive studies (Supplemental Digital Table 2) and a listing of duplicate reports (Supplemental Digital List 2) (both available at <http://links.lww.com/ACADMED/A23>).

stress disorders, trauma management, ethics, and clinical trial data abstraction to medical, dental, nursing, physician assistant, pharmacy, physical therapy, and veterinary students, and to practitioners in most of these health professions.

### Studies without comparison

We describe the 98 studies without comparison or qualitative analysis in Supplemental Digital Table 1, at <http://links.lww.com/ACADMED/A23>. Sixty of these studies (61%) reported no outcomes, while the others reported postintervention assessments of satisfaction ( $N = 25$ ), clinical reasoning ( $N = 19$ ), knowledge ( $N = 6$ ), and other skills ( $N = 5$ ). Unfortunately, without comparison with a preintervention assessment or another study group, it is difficult to make judgments regarding efficacy. However, these studies demonstrate creative approaches to specific topics, learner groups, and design challenges, and even without evidence of efficacy this information may suggest potential solutions to developers today.

### Studies using formal qualitative methods

We identified four studies that used rigorous qualitative methods (Table 1).<sup>60,88–90</sup> The methods and the research questions varied among these studies, all of which involved medical students, yet we identified several common themes. First, virtual patients have perceived advantages over other instructional approaches, including student independence, accommodation of student schedules, efficient contributions to the student's mental case library, and an unstressful learning environment. Yet, simultaneously, students believe virtual patients should not replace real patient experiences. Second, students perceive that natural case progression (data gathering, more choices and less constrained choices, and evolution in response to learner actions) and feedback are important determinants of satisfaction and engagement. Third, students identified multiple factors contributing to a case's realism, including the case material (i.e., does the script appear

authentic?), the type of case (does this case represent patients I would expect to see?), and the computer presentation (e.g., good acting in video clips, or realistic dialogue flow in virtual conversations). Finally, students generally advocated group rather than individual case completion, citing greater engagement, the requirement to defend one's choices, and the opportunity to learn both knowledge and alternative clinical approaches from one another.

### Studies with comparison: characteristics and quality

Table 2 summarizes key features of the comparative studies (see Supplemental Digital Table 2 at <http://links.lww.com/ACADMED/A23> for additional details). A total of 3,285 learners participated, including 2,115 medical students, 437 dental students, 272 nursing students, 89 physicians in postgraduate training, 34 practicing nurses, 12 physicians in practice, and 326 other learners (other allied health or mixed groups). Twenty-two of 45 studies (49%) reported skills outcomes (communication skills or clinical proficiency in a test setting), 21 (47%) reported knowledge, 15 (33%) reported clinical reasoning, and 12 (27%) reported satisfaction; none reported behaviors in practice or effects on patients.

Learners requested information from the virtual patient using menus ( $N = 24$ ; 53%), natural language ( $N = 10$ ; 22%), or both ( $N = 2$ ; 4%); we could not determine this for 9 studies. Two studies<sup>45,54</sup> compared methods of requesting information. Fourteen virtual patients (31%) used free navigation case progression, 12 (28%) used a linear pattern, 10 (22%) were branching, and one study compared free versus branching progression, while in 8 studies the case progression could not be determined. Learners collaborated in groups to complete virtual patients in 6 (13%) studies. Twenty-three virtual patients (55%) provided high interactivity, 15 (36%) provided high feedback, 8 (18%) incorporated many learning strategies, 17 (38%) provided opportunity for repetitive practice, 16 (36%) were integrated into the curriculum, and 26 (58%) reflected clinical variation in disease presentation. None of the virtual patients reflected a range of task difficulty, and none provided for individualized learning aside

**Table 1**  
**Qualitative Studies and Themes Identified in a 2009 Systematic Review of Research on Computerized Virtual Patients (VPs)**

First author, year	Participants*	Clinical topic	Virtual patient	Information sources	Analytic method	Main themes
Bryce, 1998 <sup>60</sup>	55 MS-P	Various	Menu-driven, branching	Live observation; postcase written feedback; postcase focus groups	Consensus of research team; asked students to review final themes (member check)	<p>Advantages of VP:</p> <ul style="list-style-type: none"> <li>● Student performed all actions of primary caregiver (not just bystander on team)</li> <li>● Student independence</li> <li>● Developed reasoning skills: increased mental pool of "experienced" cases, explicit reasoning path</li> <li>● Data-gathering opportunities</li> <li>● Feedback</li> <li>● Accommodated student schedule</li> <li>● Could look up information while working on case</li> <li>● Less stressful; could think clearly, not embarrassed or intimidated</li> <li>● Could experiment with patient</li> <li>● Helped identify knowledge gaps</li> <li>● Other themes</li> <li>● VP should not replace real patients</li> <li>● Group work generally preferred (draws on others' knowledge, must justify decisions) although some groups dysfunctional</li> <li>● Some students treated encounter as real</li> <li>● Others used encounter to experiment (try actions they would never do in real life)</li> </ul>
Bearman, 2003 <sup>88</sup>	12 MS-P	Chest pain/communication skills	Menu-driven; problem-solving (free navigation) and narrative (branching) formats	Unstructured interviews of purposive student sample	Phenomenology; interviews transcribed, themes extracted (bracketing, horizontalization, imaginative variation) and common elements and differences identified; asked students to review final themes (member check)	<ul style="list-style-type: none"> <li>● Students retained identify while working on case even as they assumed physician role</li> <li>● VP reflected a real type of patient, but lacked complexity</li> <li>● Problem-solving format was more realistic, but also more frustrating because actions had no impact (patient did not evolve over time)</li> </ul>
Bergin, 2003 <sup>89</sup>	24 MS-P; 28 MS-C	Fever of unknown origin	Natural-language, free navigation; some students also completed paper patient cases	Live observation; focus groups; 1-on-1 interviews; session videotapes	"Analyzed by evaluation team"	<ul style="list-style-type: none"> <li>● VP should not replace real patients</li> <li>● Group work preferred (learn knowledge and approaches to case; more engaged)</li> <li>● Two forms of realism: realistic actor, realistic case material</li> <li>● Feedback essential</li> <li>● VP more realistic and engaging than paper</li> </ul>
Mallott, 2005 <sup>90</sup>	80 MS-C	Various surgery, gynecology, medicine	Menu-driven, linear	Electronic log (differential diagnosis, time to complete, amount of trial and error, congruence with mentor solution); postcase interview	Iterative process involving research team and students to arrive at consensus on dominant themes	<p>VP should:</p> <ul style="list-style-type: none"> <li>● Accommodate trial and error</li> <li>● Allow unconstrained action (queries, interventions)</li> <li>● Change in response to user input</li> <li>● Permit student to advance the clock</li> <li>● Allow both menu and natural language</li> <li>● Permit unlimited time</li> <li>● Allow students to record key information and potential diagnoses</li> <li>● Match difficulty level to the individual</li> <li>● Compare learner responses with a mentor's approach</li> </ul>

\* MS-P indicates preclinical medical student; MS-C, clinical-years medical student.

Table 2

**Description of 43 Comparative Studies Included in a 2009 Systematic Review of Research on Computerized Virtual Patients\***

First author, year	Participants <sup>†</sup>	Study design	Clinical topic	Comparison <sup>‡</sup>	Outcomes <sup>§</sup>
Bitzer, 1966 <sup>43</sup>	NS; N = 14	Pretest–posttest, 2 groups	Myocardial infarction	T	K
Cassidy, 1972 <sup>44</sup>	DS; N = 16	Pretest–posttest, 1 groups	Toothache		C
Mullaney, 1976 <sup>45</sup>	DS; N = 147	Posttest-only, 3 groups	Endodontics	1. VP/CAI 2. ST	R, C
Murray, 1977 <sup>46</sup>	MS; N = 22	Pretest–posttest, 2 groups	General medicine topics	NI	K, C
Schleutermann, 1983 <sup>47</sup>	NPS; N = 12	Posttest-only, 2 groups	Ambulatory medicine	P	R, S
Dale, 1986 <sup>48</sup>	DS; N = 24	Posttest-only, 2 groups	Clinical endodontics	VP/CAI	R, C
Garrett, 1986 <sup>49</sup>	R; N = 14	Pretest–posttest, 1 groups	Lung cancer		C
Krahn, 1986 <sup>50</sup>	MS; N = 79	Pretest–posttest, 2 groups	Acid–base disturbances	VP/CAI	K
Sandoval, 1987 <sup>51</sup>	DS; N = 105	Pretest–posttest, 8 groups <sup>¶</sup>	Endodontics	1. ST 2. P 3. VP/CAI	R, C
Harless, 1990 <sup>52</sup>	MS; N = 80	Pretest–posttest, 1 groups	GI bleed, obesity		K
Lyon, 1990 <sup>53</sup>	MS; N = 176	Posttest-only, 2 groups	Chest pain, anemia	VP/CAI	K
Friedman, 1991 <sup>54</sup>	MS; N = 80	Posttest-only, 3 groups	Various medical topics	1. VP/CAI 2. VP/CAI	R, K
Lowdermilk, 1991 <sup>55</sup>	NS; N = 64	Pretest–posttest, 2 groups	Various clinical nursing topics	CT	C
Lyon, 1991 <sup>56</sup>	MS; N = 169	Pretest–posttest, 2 groups	Chest pain, anemia	P	K
Weverling, 1996 <sup>57</sup>	MS; N = 103	Posttest-only, 2 groups	Neurology	NI	K
Johnson, 1997 <sup>58</sup>	DH; N = 58	Pretest–posttest, 3 groups	Geriatric dental hygiene	1. T 2. NI	C
Kinney, 1997 <sup>59</sup>	PTS; N = 10	Pretest–posttest, 2 groups	Carpal tunnel syndrome	T	K
Bryce, 1998 <sup>60</sup>	MS; N = 110	Posttest-only, 2 groups	Not specified	NI	K**
Schwid, 1999 <sup>61</sup>	R, PP; N = 45	Posttest-only, 2 groups	ACLS	P	S
Fleetwood, 2000 <sup>62</sup>	MS; N = 173	Posttest-only, 2 groups	Medical ethics	T	K, S
Bearman, 2001 <sup>63</sup>	MS; N = 255	Posttest-only, 3 groups	Communication skills	1. VP/CAI 2. NI	S
Schwid, 2001 <sup>64</sup>	R; N = 31	Posttest-only, 2 groups	Anesthesia emergencies	P	S
Buysse, 2002 <sup>65</sup>	MS; N = 82	Posttest-only, 2 groups	Ambulatory medicine	VP/CAI	R
Chaikoolvatana, 2003 <sup>66</sup>	PS; N = 79	Posttest-only, 2 groups	Pharmacy care plans	SP	S
Kumta, 2003 <sup>67</sup>	MS; N = 163	Posttest-only, 2 groups	Orthopedics	CT, T	S
Hayes-Roth, 2004 <sup>68</sup>	MS, NS; N = 31	Posttest-only, 2 groups	Motivational interviewing	VP/CAI	R, S
Schitteck Janda, 2004 <sup>69</sup>	DS; N = 50	Posttest-only, 2 groups	Interviewing skills	NI	S
Dickerson, 2006 <sup>70</sup>	MS; N = 17	Posttest-only, 2 groups	Abdominal pain	VP/CAI	R, S
Ferguson, 2006 <sup>71</sup>	R; N = 30	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Raij, 2006 <sup>72</sup>	MS; N = 24	Posttest-only, 2 groups	Abdominal pain	SP	R, C, S
Thompson, 2006 <sup>73</sup>	MS; N = 96	Posttest-only, 2 groups	Cardiology and bioterrorism	VP/CAI	K, C
Triola, 2006 <sup>74</sup>	NP, PP, psychologist, other; N = 55	Pretest–posttest, 2 groups	Distress disorders	SP	R, C
Turner, 2006 <sup>75</sup>	MS; N = 30	Posttest-only, 2 groups	Abdominal pain, headache	SP	R, C, S
Wahlgren, 2006 <sup>76</sup>	MS; N = 116	Posttest-only, 2 groups	Dermatology, venereology	NI	K
Deladisma, 2007 <sup>77</sup>	MS; N = 84	Posttest-only, 2 groups	Abdominal pain	SP	S
Kleinert, 2007 <sup>78</sup>	PAS; N = 42	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Kleinert, 2007 <sup>79</sup>	DS; N = 51	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Raij, 2007 <sup>80</sup>	MS; N = 58	Posttest-only, 2 groups	Abdominal pain	SP	C, S
Sanders, 2007 <sup>81</sup>	NS; N = 98	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Sijstermans, 2007 <sup>82</sup>	MS; N = 134	Pretest–posttest, 1 groups	Interphysician communication		C
Vash, 2007 <sup>83</sup>	MS; N = 48	Posttest-only, 2 groups	Abdominal pain	CT	K, C

(Continues)

Table 2

(Continued)

First author, year	Participants <sup>†</sup>	Study design	Clinical topic	Comparison <sup>*</sup>	Outcomes <sup>§</sup>
Boyd, 2008 <sup>84</sup>	PAS, NS, R; N = 101	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Sanders, 2008 <sup>85</sup>	DS; N = 44	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Sanders, 2008 <sup>86</sup>	NPS; N = 35	Pretest–posttest, 1 groups	Developmental disabilities		K, S
Youngblood, 2008 <sup>87</sup>	MS, R; N = 30	Pretest–posttest, 2 groups	Trauma management	HPS	R, S

\* Additional details are provided in Supplemental Digital Table 3, at <http://links.lww.com/ACADMED/A23>.

<sup>†</sup> NS indicates nursing student; DS, dental student; MS, medical student; NPS, nurse practitioner student; R, resident (physician in postgraduate training); DH, dental hygienists; PTS, physical therapy student; PP, physician in practice; NP, nurses in practice; PAS, physician assistant student.

<sup>\*</sup> Comparison intervention: CT indicates clinical teaching; HPS, human patient simulator (manikin); P, paper (handout, textbook, or latent image cases); NI, no-intervention control group; SP, standardized patient; ST, slide-tape; T, traditional (typically lecture); VP/CAI, virtual patient/computer-assisted instruction. Blank cells indicate no comparison group (i.e., one-group pretest–posttest study).

<sup>§</sup> Outcomes reported in text: R indicates reaction (satisfaction); K, knowledge; C, clinical reasoning; S, skills.

<sup>¶</sup> Sandoval<sup>51</sup> compared slide-tape instruction, latent image paper cases, two virtual patients, and four combinations of these interventions (eight groups). We abstracted information on the four main interventions.

\*\* Insufficient data reported to determine effect size.

from branching scenarios based on learner choices.

Table 3 summarizes the methodological quality of the comparative studies. Twenty-one studies (62%) were randomized. Twenty-one of 21 (100%) knowledge assessments, 14 of 15 (93%) clinical reasoning assessments, and 17 of 22 (77%) skill assessments used objective measures. Two of 12 studies (17%) compared course completion rates as a measure of satisfaction; all other satisfaction measures were self-reported. Five (24%) studies assessing knowledge, 5 (33%) assessing clinical reasoning, 6 (27%) assessing skills, and 4 (33%) assessing satisfaction lost more than 25% of participants from time of enrollment or failed to report follow-up. Quality scores (6 points indicating highest quality) ranged from 0 to 6, with mean (SD) 3.2 (1.4) and median 3.

### Comparisons with no intervention

Eighteen studies (1,359 participants) reported comparison with a preintervention assessment or a no-intervention control group. Of these, 11 reported knowledge outcomes (Figure 2), with a pooled ES of 0.94 (95% CI, 0.69–1.19,  $P < .001$ ). Because ESs  $> 0.8$  are considered large,<sup>42</sup> this suggests that virtual patient interventions are associated with substantial knowledge gains. However, we also found large inconsistency among studies, with ESs ranging from 0.27 to 2.07 and  $I^2 = 81\%$ . An asymmetric funnel plot suggested possible publication bias. Assuming this

asymmetry reflects publication bias, trim and fill analyses provided a revised pooled ES of 0.90 (95% CI, 0.65–1.15).

For the five studies reporting clinical reasoning outcomes, the pooled ES was large (0.80 [95% CI, 0.52–1.08],  $P < .001$ ), with moderate inconsistency ( $I^2 = 46\%$ ). The funnel plot appeared symmetric.

Nine studies reported skill outcomes, with a large pooled ES of 0.90 (0.61–1.19,  $P < .001$ ) and large inconsistency ( $I^2 = 82\%$ ). The funnel plot was asymmetric. Again assuming that this reflects bias, trim and fill analyses yielded a revised pooled ES of 0.79 (95% CI, 0.48–1.10).

In planned subgroup analyses, we found no statistically significant interactions with virtual patient design features of interactivity, feedback, number of instructional strategies, or time spent learning (see Supplemental Digital Table 3, at <http://links.lww.com/ACADMED/A23>). We found no significant interaction with blinding or overall quality score, but we did find a significant interaction with number of groups; namely, two-group studies demonstrated a smaller pooled ES (0.49) than one-group pretest–posttest studies (0.92;  $P_{\text{interaction}} = .015$ ). We obtained virtually identical results for all outcomes in sensitivity analyses excluding studies published before 1991.

### Comparisons with noncomputer interventions

Twenty articles reported 21 studies (1,546 participants) comparing virtual patients

with various noncomputer interventions, including traditional instruction (typically lecture), standardized patients, paper instruction (handouts, textbooks, or latent-image paper cases), slide-tape instruction, routine clinical activities, and training with a physiologically responsive manikin. One study<sup>51</sup> compared a virtual patient with both latent image and slide-tape instruction. Because these comparisons are not independent, we selected one—slide-tape instruction—for reported meta-analyses. However, sensitivity analyses substituting the latent image data yielded virtually identical results.

For the five studies reporting knowledge outcomes (Figure 3), the pooled ES was 0.06 (95% CI,  $-0.14$  to  $0.25$ ;  $P = .56$ ) with  $I^2 = 0$ , with positive numbers favoring the virtual patient intervention. Because ESs  $< 0.2$  are considered small effects,<sup>42</sup> this suggests that virtual patients are associated with rather negligible differences in knowledge outcomes compared with other active instructional activities.

The pooled ES for the 10 studies reporting reasoning outcomes was  $-0.004$  (95% CI,  $-0.30$  to  $0.29$ ;  $P = .98$ ) with  $I^2 = 70\%$ . Eleven studies reported skill outcomes, with a pooled ES of 0.10 (95% CI,  $-0.21$  to  $0.42$ ;  $P = .52$ ) and  $I^2 = 84\%$ . As with knowledge, this suggests small and statistically nonsignificant associations between use of virtual patients and other instructional methods for reasoning or skill outcomes. Finally,

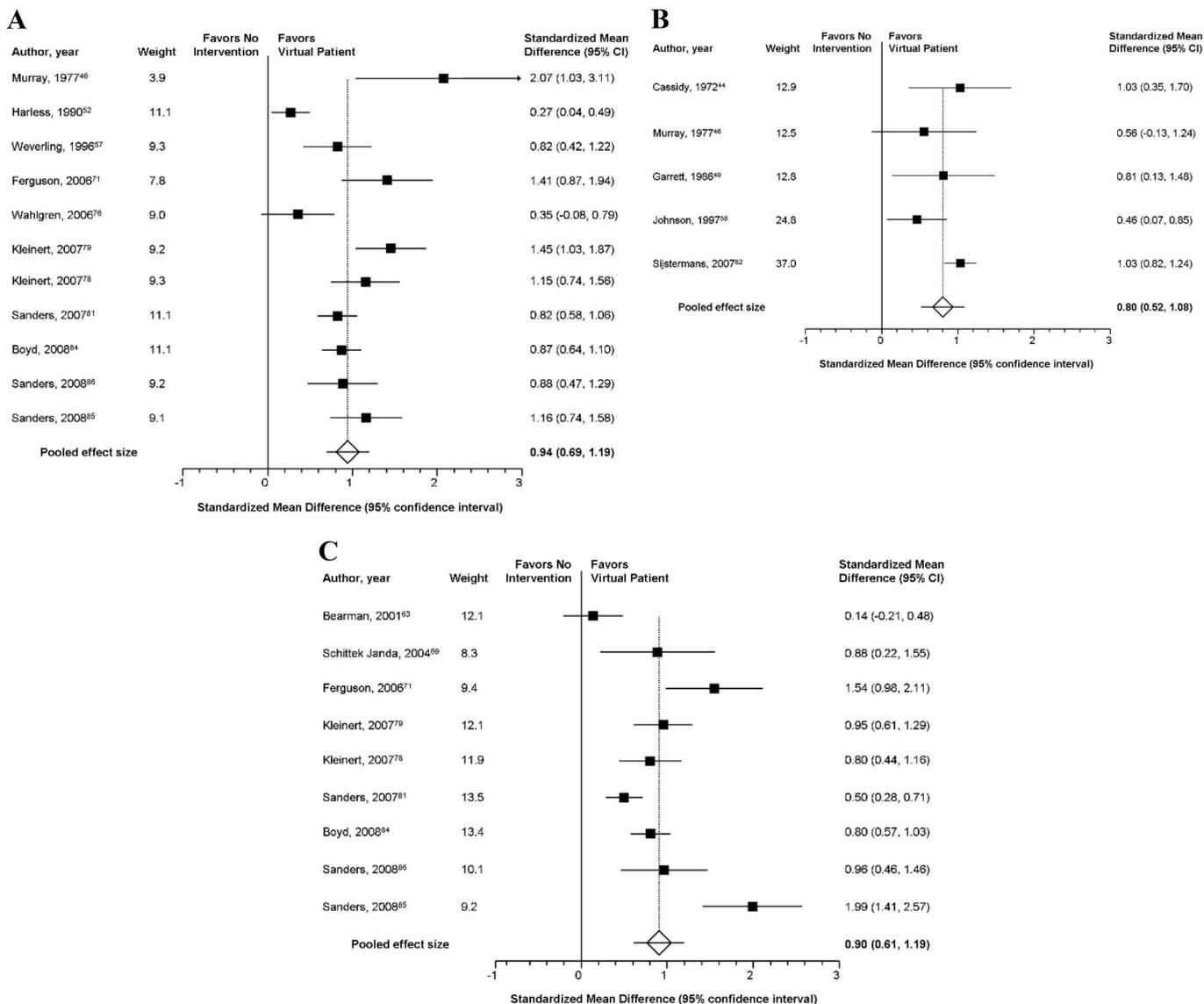
Table 3

**Quality of 43 Comparative Studies Included in a 2009 Systematic Review of Research on Computerized Virtual Patients\***

First author, year	Representative intervention group	Comparison group selected from same community	Comparability of cohorts	Blinded outcome assessment <sup>†</sup>	Follow-up adequate <sup>†</sup>
Bitzer, 1966 <sup>43</sup>	Yes	Yes		Yes	Yes
Cassidy, 1972 <sup>44</sup>	Yes			Yes	Yes
Mullaney, 1976 <sup>45</sup>		Yes	Randomized	Yes	Yes
Murray, 1977 <sup>46</sup>		Yes		Yes	
Schleutermann, 1983 <sup>47</sup>		Yes	Randomized	Yes	Yes
Dale, 1986 <sup>48</sup>		Yes			Yes
Garrett, 1986 <sup>49</sup>				Yes	Yes
Krahn, 1986 <sup>50</sup>	Yes	Yes		Yes	Yes
Sandoval, 1987 <sup>51</sup>	Yes	Yes		Yes	Yes
Harless, 1990 <sup>52</sup>	Yes			Yes	
Lyon, 1990 <sup>53</sup>					
Friedman, 1991 <sup>54</sup>	Yes	Yes	Randomized, allocation concealed	Yes	Yes
Lowdermilk, 1991 <sup>55</sup>	Yes	Yes	Randomized	Yes	
Lyon, 1991 <sup>56</sup>			Controlled for learning outcome		
Weverling, 1996 <sup>57</sup>	Yes	Yes	Randomized, allocation concealed	Yes	Yes
Johnson, 1997 <sup>58</sup>				Yes	
Kinney, 1997 <sup>59</sup>		Yes	Randomized	Yes	Yes
Bryce, 1998 <sup>60</sup>		Yes	Randomized	Yes	
Schwid, 1999 <sup>61</sup>		Yes	Randomized	Yes	Yes
Fleetwood, 2000 <sup>62</sup>	Yes	Yes		Yes	Yes
Bearman, 2001 <sup>63</sup>	Yes	Yes	Randomized	Yes	Yes
Schwid, 2001 <sup>64</sup>		Yes	Randomized	Yes	Yes
Buysse, 2002 <sup>65</sup>	Yes	Yes	Randomized		Yes
Chaikoolvatana, 2003 <sup>66</sup>	Yes	Yes			
Kumta, 2003 <sup>67</sup>		Yes	Randomized		
Hayes-Roth, 2004 <sup>68</sup>		Yes	Controlled for learning outcome, other	Yes	
Schittek Janda, 2004 <sup>69</sup>	Yes	Yes	Randomized, allocation concealed	Yes	Yes
Dickerson, 2006 <sup>70</sup>		Yes	Randomized		
Ferguson, 2006 <sup>71</sup>	Yes			Yes	Yes
Raij, 2006 <sup>72</sup>					
Thompson, 2006 <sup>73</sup>		Yes	Randomized	Yes	Yes
Triola, 2006 <sup>74</sup>		Yes	Randomized	Yes	Yes
Turner, 2006 <sup>75</sup>		Yes	Randomized	Yes	Yes
Wahlgren, 2006 <sup>76</sup>	Yes	Yes	Randomized	Yes	Yes
Deladisma, 2007 <sup>77</sup>		Yes	Randomized		Yes
Kleinert, 2007 <sup>78</sup>	Yes			Yes	Yes
Kleinert, 2007 <sup>79</sup>	Yes			Yes	Yes
Raij, 2007 <sup>80</sup>	Yes	Yes			
Sanders, 2007 <sup>81</sup>	Yes			Yes	Yes
Sijstermans, 2007 <sup>82</sup>					Yes
Vash, 2007 <sup>83</sup>		Yes	Randomized, allocation concealed	Yes	Yes
Boyd, 2008 <sup>84</sup>	Yes			Yes	Yes
Sanders, 2008 <sup>85</sup>	Yes			Yes	Yes
Sanders, 2008 <sup>86</sup>	Yes			Yes	Yes
Youngblood, 2008 <sup>87</sup>		Yes	Randomized		Yes

\* Quality was assessed using a modification of the Newcastle–Ottawa Scale.<sup>31</sup> Each study could receive up to six points (maximum two points for comparability of cohorts; one point for other criteria). See Cook et al<sup>20</sup> for details.

<sup>†</sup> Blinding and completeness of follow-up ( $\geq 75\%$  follow up or description provided of those lost) are reported as Yes if this was true for any reported outcome.



**Figure 2** Random-effects meta-analysis of virtual patients in comparison with no intervention. Boxes represent the standardized mean difference (Hedges g effect size), and bars represent the 95% confidence interval (95% CI). Open diamonds represent pooled overall estimates. **Panel A:** Knowledge outcomes.  $I^2 = 81\%$  for this analysis. **Panel B:** Clinical reasoning outcomes.  $I^2 = 46\%$  for this analysis. **Panel C:** Skill outcomes.  $I^2 = 82\%$  for this analysis.

the eight studies evaluating satisfaction outcomes yielded a pooled ES of  $-0.17$  (95% CI,  $-0.57$  to  $0.24$ ;  $P = .42$ ) with  $I^2 = 71\%$ .

Subgroup analyses exploring associations between methodological quality or virtual patient design features and performance revealed no statistically significant interactions (see Supplemental Digital Table 3 at <http://links.lww.com/ACADMED/A23>). Funnel plots and the Egger asymmetry test did not suggest publication bias for any outcomes. Sensitivity analyses excluding studies published before 1991 yielded almost identical results for all outcomes.

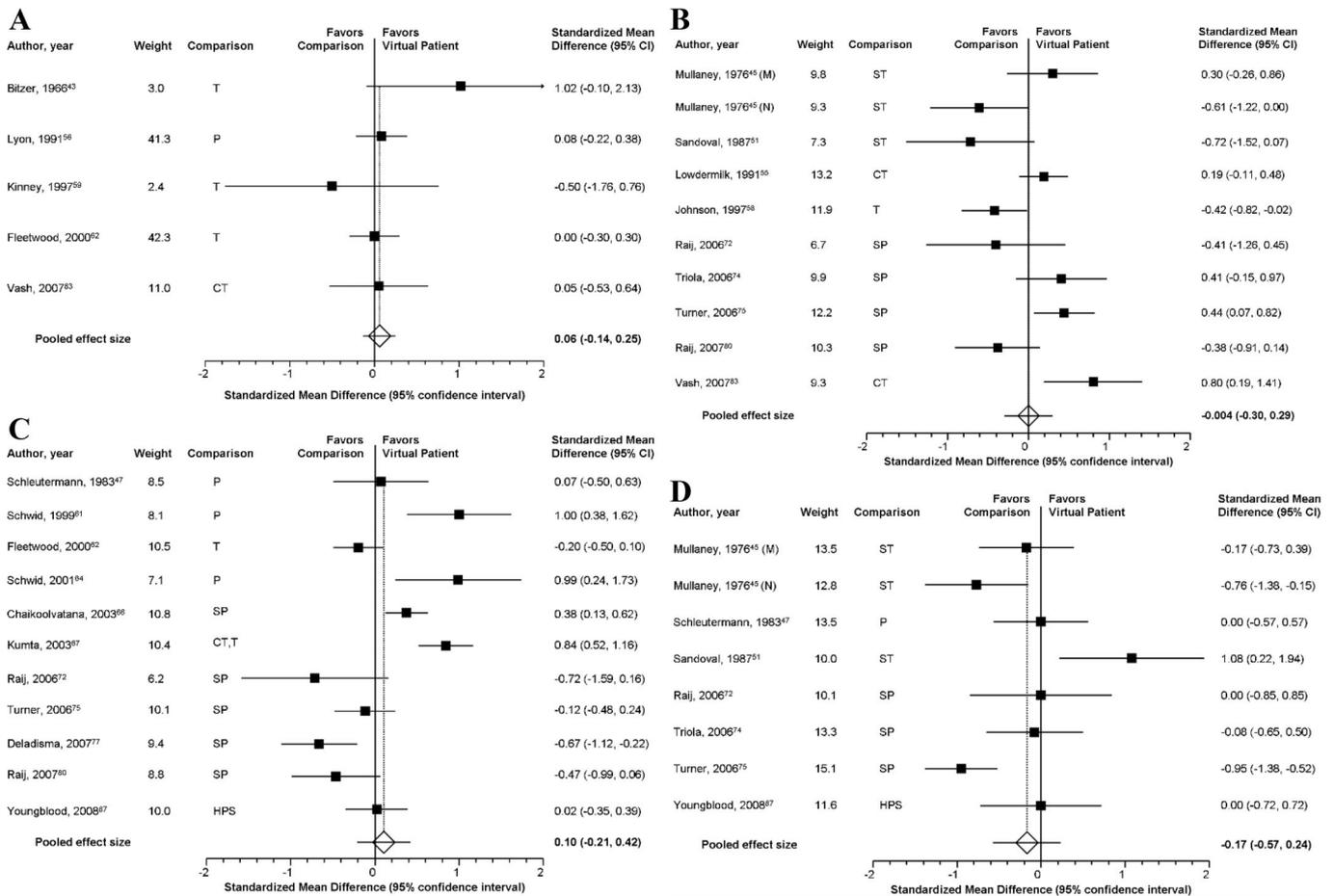
### Comparisons between virtual patient designs

Comparisons between virtual patient formats can illuminate how different virtual patient design features affect learning outcomes. Eleven studies took this approach, comparing one virtual patient with another.<sup>45,48,50,51,53,54,63,65,68,70,73</sup>

Because the differences between virtual patients varied substantially for each study, we could not perform a quantitative synthesis, so we present a narrative synthesis instead. Because study designs and statistical tests varied, we report ES and sample size (which in some substudies is smaller than that reported

in Table 2) rather than tests of statistical significance. Space limitations prohibit a full description of each method and context, and interested readers may wish to consult the original studies for additional details.

Four studies explored different methods of information exchange. One of these<sup>45</sup> compared a natural language user interface, in which learners typed questions to elicit information from the virtual patient, with a menu-driven interface. The data do not permit direct estimation of an ES for clinical reasoning outcomes; however, comparison with a common slide-tape intervention (in



**Figure 3** Random-effects meta-analysis of virtual patients in comparison with other noncomputer instructional activities. Boxes represent the standardized mean difference (Hedges g effect size), and bars represent the 95% confidence interval (95% CI). Open diamonds represent pooled overall estimates. CT indicates clinical teaching; HPS, human patient simulator (manikin); P, paper (handout, textbook, or latent image cases); SP, standardized patient; ST, slide-tape; T, traditional (typically lecture). **Panel A:** Knowledge outcomes.  $I^2 = 0.0\%$  for this analysis. **Panel B:** Clinical reasoning outcomes.  $I^2 = 70\%$  for this analysis. Mullaney et al<sup>45</sup> reported two separate studies comparing a slide-tape intervention with a menu-driven virtual patient (M) and a virtual patient using free text (natural language) input (N). **Panel C:** Skill outcomes.  $I^2 = 84\%$  for this analysis. **Panel D:** Satisfaction outcomes.  $I^2 = 71\%$  for this analysis. Mullaney et al<sup>45</sup> reported two studies comparing a slide-tape intervention with a menu-driven virtual patient (M) and a virtual patient using free text (natural language) input (N).

different study years) revealed outcomes favoring the menu format (ES 0.30, N = 49) and opposing the natural language approach (ES -0.61, N = 44). The menu-driven format was also associated with greater satisfaction (ES 0.32). In another study,<sup>70</sup> the virtual patient spoke using either synthesized or prerecorded speech. This small randomized trial (N = 17) found differences modest in magnitude but not statistically significant, with higher voice clarity ratings (ES 0.37) and higher performance scores (0.37) for the recorded speech format. Two studies<sup>48,51</sup> compared text-only virtual patients with virtual patients enhanced with video. Results varied between studies, with one study<sup>48</sup> showing higher satisfaction (ES 0.47) and reasoning (ES 0.23, N = 24) outcomes for the text-only format, and the other

study<sup>51</sup> showing no preference (satisfaction ES 0.0, N = 26) and slightly worse reasoning (ES -0.18) for the text-only format.

Five studies explored different instructional methods. One randomized trial<sup>54</sup> found that a menu-driven virtual patient with advance organizers and detailed feedback improved participants' knowledge more than natural language lower-feedback formats designed to encourage hypothesis generation (ES 0.79, N = 48) or emulate a real clinical encounter (ES 0.82, N = 50), although learners reported lower satisfaction (ES  $\leq -0.54$ ). Advance organizers also improved knowledge in another randomized study (ES 0.60, N = 79).<sup>50</sup> Requiring learners to contrast each new virtual patient case with prior cases led to

small knowledge gains in a third randomized trial (ES 0.30, N = 50).<sup>73</sup> A fourth randomized study<sup>68</sup> evaluated an interactive virtual patient system in which a virtual coach required repetition until learners demonstrated mastery. In comparison with static Web pages with similar content, this system enhanced participants' interviewing skills (ES 1.47, N = 22) but had similar satisfaction scores (ES 0.0). Finally, a historical-control study<sup>53</sup> compared an early virtual patient system with a later version (the following academic year) enhanced with more cases, more features, and a mandatory completion requirement. Although student classroom time decreased 12 hours in the second year, knowledge scores were similar for both groups (ES 0.08, N = 176).

Two randomized studies explored different ways to structure the virtual patient interaction. The first, described above, found improved knowledge but decreased satisfaction for structured, educationally enriched virtual patients compared with realistic, unstructured cases.<sup>54</sup> The other study found similar communication skills following use of an unstructured problem-solving format and a format structured to emphasize temporal relationships (ES 0.12, N = 157),<sup>63</sup> although a phenomenological qualitative study found that learners established better rapport with the narrative patient.<sup>88</sup>

Finally, a study found that imposing a two-hour time limit lowered the rate of case completion (ES 2.13, N = 82).<sup>65</sup>

## Discussion

We found that virtual patients, in comparison with no intervention, are consistently associated with higher learning outcomes. Pooled ESs were large ( $\geq 0.80$ )<sup>42</sup> for outcomes of knowledge, clinical reasoning, and other skills, and CIs excluded small effects ( $< 0.5$ ). However, the magnitude of effect varied for individual studies (large inconsistency), and subgroup analyses exploring differences in virtual patient designs largely failed to explain this variation. By contrast, the pooled ESs for studies comparing virtual patients with noncomputer interventions were small ( $-0.17$  to  $0.10$ ) and nonsignificant (CIs encompassing zero [no effect]). CIs excluded moderate effects ( $\geq 0.5$ ) but could not exclude small effects ( $0.2$  to  $0.5$ ). Once again, inconsistency (heterogeneity) among studies was large, and subgroup analyses did little to explain these inconsistencies.

Although the above subgroup analyses did not answer our question regarding the effectiveness of different virtual patient designs, comparisons between virtual patient formats address this issue. For example, mastery learning, advance organizers, enhanced feedback, and explicitly contrasting cases improved learning outcomes in randomized trials, with ESs ranging  $0.29$  to  $1.47$ . Variations in virtual patient structure and the method of information exchange were also associated with differences in learning outcomes. Qualitative research studies further suggest that natural case

evolution and working as groups are important. These findings suggest that at least some of the inconsistency noted above arises from differences in interventions.

Subgroup analyses of no-intervention-comparison studies revealed a significant interaction with study design, with two-group studies demonstrating smaller pooled ESs and somewhat lower inconsistency than one-group studies. It makes sense that studies with a comparison group, which helps control for maturation and learning outside the intervention, would show smaller effects than single-group studies. However, these findings could also be due to chance or to other between-study differences such as variation in virtual patient design, concurrent nonvirtual patient learning opportunities, and the sensitivity of the outcome measure. By contrast, we found no statistically significant interactions with other quality measures for no intervention or media-comparative studies.

## Limitations and strengths

As in any review, the inferences we draw are limited by the quantity and quality of available studies. Many reports failed to clearly describe key features of the context, instructional design, or outcomes. Fewer than half the comparative studies were randomized, and most studies had other important methodological limitations. The modest number of studies and participants limits the precision of our meta-analysis results and the power of our subgroup analyses. The age of some studies makes them of questionable relevance, but excluding older studies did not appreciably alter the results. We found large inconsistency among studies, and statistical pooling cannot account for all potentially important differences in learner groups, clinical topics, interventions, study designs, and outcome measures.<sup>93</sup> However, because all no-intervention-comparison studies favored virtual patients, this heterogeneity suggests that virtual patients may be effective across a broad range of learners and topics. Because virtual patients are designed for health professions training, we did not include studies from non-health-related fields. Finally, of necessity, we abstracted information on only a few virtual patient design features. Although we selected these features after considering numerous

possibilities<sup>19</sup> and evidence from related fields,<sup>15</sup> we still might have missed important features.

Our review also has several strengths, including a timely and important question; a systematic literature search aided by an experienced reference librarian, including multiple databases and supplemented by hand searches; explicit and reproducible inclusion criteria encompassing a broad range of learners, outcomes, and study designs; duplicate, independent, and reproducible data abstraction; rigorous coding of methodological quality; and focused analyses. We reviewed in detail both quantitative comparative and qualitative studies and summarized many descriptive studies including several non-English reports (see Supplemental Digital Table 1, at <http://links.lww.com/ACADMED/A23>). We used funnel plots to assess for publication bias, and although this method is limited in the presence of large inconsistency,<sup>38</sup> it did not suggest that publication bias substantially affected our conclusions.

## Comparison with previous reviews

To our knowledge, this is the first systematic review to address the topic of virtual patients in health professions education. A recent narrative review<sup>19</sup> identified a number of important questions regarding virtual patients, but it selectively included studies and did not provide a quantitative synthesis of outcomes. Similar to the present study, a meta-analysis of laparoscopic surgery simulation<sup>16</sup> found improved outcomes for simulation training compared with no training, as did systematic reviews of surgical simulation in general<sup>17</sup> and of colonoscopy and laparoscopic cholecystectomy simulation.<sup>18</sup> Another systematic review suggested that feedback, curricular integration, and multiple learning strategies are essential features of simulation<sup>15</sup>; we cannot corroborate or refute these conclusions. Our findings of large ESs for comparisons with no intervention and small ESs for comparisons with other active interventions are consistent with a recent meta-analysis of Internet-based instruction.<sup>20</sup>

## Implications

The use of virtual patients as learning tools is associated with improved

outcomes in comparison with no intervention for medical students, dental students, nursing students, and a variety of other health professionals across a range of clinical topics. Evidence does not indicate superiority of virtual patients over other training methods, but allowing for the uncertainty of the CIs and imperfections of the outcome measures, they may be noninferior in some instances. Inasmuch as virtual patients resolve logistic barriers<sup>13,14,94</sup> or provide unquantified advantages (such as those identified in the qualitative studies or predicted by education theories), they may warrant use to enhance cognitive clinical skills among student and practicing health professionals.

The virtual patients we identified varied widely in their design, implementation, and effectiveness. Unfortunately, available evidence answers only in part our question regarding what virtual patient design variations lead to improved learning outcomes. Subgroup analyses failed to identify significant interactions involving instructional designs, but between-study (rather than within-study) comparisons are an inefficient research method.<sup>95</sup> By contrast, direct comparisons of two virtual patient designs were few but generally supported theories predicting that cognitive interactivity, learning to mastery, and feedback yield better outcomes.

We believe that theory-based comparisons between different virtual patient designs, and rigorous qualitative studies, will clarify how to effectively use virtual patients for training health professionals. Frameworks such as multimedia learning,<sup>96,97</sup> analytical and nonanalytical reasoning,<sup>19</sup> deliberate practice,<sup>98</sup> and formative feedback<sup>99,100</sup> may be useful. The associations found in several studies between changes intended to make the virtual patient more realistic and neutral or negative outcomes raise questions regarding for whom, in what contexts, and for what outcomes greater realism is beneficial.<sup>101</sup> Most research to-date has involved students; the role of virtual patients in postgraduate and continuing education requires further study. Research outcomes have largely

focused on short-term knowledge, clinical reasoning, and other skills. Perhaps new measures (e.g., different clinical reasoning assessments<sup>19</sup>) or different outcomes (e.g., decision-making behaviors, health care costs, or medical errors) would more closely align with the long-term objectives of using virtual patients. Finally, we hope that future researchers can avoid the weaknesses of previous research by designing studies that minimize bias, achieve appropriate power, and avoid confounding.<sup>102</sup>

*Funding/Support:* This work was supported by intramural funds and by a Commissioned Review Award from the Society of Directors of Research in Medical Education. The funding sources for this study played no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation of the manuscript. The funding sources did not review the manuscript.

*Other disclosures:* None.

*Ethical approval:* As no human subjects were involved, ethical approval was not required.

*Previous presentations:* Portions of this work were presented in symposia at the 2009 meetings of the Association for Medical Education in Europe (Málaga, Spain) and Association of American Medical Colleges (Boston, Massachusetts), and as an abstract at the 2010 Annual International Meeting on Simulation in Healthcare (Phoenix, Arizona).

## References

- Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Ann Intern Med.* 2006; 145:488–496.
- Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165:1493–1499.
- Singh H, Thomas EJ, Petersen LA, Studdert DM. Medical errors involving trainees: A study of closed malpractice claims from 5 insurers. *Arch Intern Med.* 2007;167:2030–2036.
- Newman-Toker DE, Pronovost PJ. Diagnostic errors—the next frontier for patient safety. *JAMA.* 2009;301:1060–1062.
- Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency department: A study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.* 2007;49:196–205.
- Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Acad Med.* 1998;73(10 suppl):S1–S5.
- Norman G, Dore K, Krebs J, Neville AJ. The power of the plural: Effect of conceptual analogies on successful transfer. *Acad Med.* 2007;82(10 suppl):S16–S18.

- Reed DA, Levine RB, Miller RG, et al. Effect of residency duty-hour limits: Views of key clinical faculty. *Arch Intern Med.* 2007;167: 1487–1492.
- Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: An ethical imperative. *Acad Med.* 2003;78:783–788.
- Effective Use of Educational Technology in Medical Education: Summary Report of the 2006 AAMC Colloquium on Educational Technology. Washington, DC: Association of American Medical Colleges; 2007.
- McGee JB, Neill J, Goldman L, Casey E. Using multimedia virtual patients to enhance the clinical curriculum for medical students. *Stud Health Technol Inform.* 1998; 52 pt 2:732–735.
- Voelker R. Virtual patients help medical students link basic science with clinical care. *JAMA.* 2003;290:1700–1701.
- Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. *Acad Med.* 2007;82:446–451.
- Ellaway R, Poulton T, Fors U, McGee JB, Albright S. Building a virtual patient commons. *Med Teach.* 2008;30:170–174.
- Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Med Teach.* 2005;27: 10–28.
- Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg.* 2008;95: 1088–1097.
- Sutherland LM, Middleton PF, Anthony A, et al. Surgical simulation: A systematic review. *Ann Surg.* 2006;243:291–300.
- Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. *Ann Surg.* 2008;248: 166–179.
- Cook DA, Triola MM. Virtual patients: A critical literature review and proposed next steps. *Med Educ.* 2009;43:303–311.
- Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: A meta-analysis. *JAMA.* 2008;300:1181–1196.
- Cohen PA, Dacanay LD. Computer-based instruction and health professions education: A meta-analysis of outcomes. *Eval Health Prof.* 1992;15:259–281.
- Greenhalgh T. Computer assisted learning in undergraduate medical education. *BMJ.* 2001;322:40–44.
- Lewis MJ, Davies R, Jenkins D, Tait MI. A review of evaluative studies of computer-based learning in nursing education. *Nurse Educ Today.* 2001;21:26–37.
- Chumley-Jones HS, Dobbie A, Alford CL. Web-based learning: Sound educational method or hype? A review of the evaluation literature. *Acad Med.* 2002;77(10 suppl): S86–S93.
- Rosenberg H, Grad HA, Matear DW. The effectiveness of computer-aided, self-instructional programs in dental education:

- A systematic review of the literature. *J Dent Educ.* 2003;67:524–532.
- 26 Welk A, Splieth C, Wierinck E, Gilpatrick RO, Meyer G. Computer-assisted learning and simulation systems in dentistry—a challenge to society. *Int J Comput Dent.* 2006;9:253–265.
  - 27 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Quality of Reporting of Meta-analyses.* *Lancet.* 1999;354:1896–1900.
  - 28 Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA.* 2000;283:2008–2012.
  - 29 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann Intern Med.* 2009;151:264–269.
  - 30 Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
  - 31 Wells GA, Shea B, O'Connell D, et al. The Newcastle–Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.htm](http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm). Accessed May 12, 2010.
  - 32 Borenstein M. Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC, eds. *The Handbook of Research Synthesis.* 2nd ed. New York, NY: Russell Sage Foundation; 2009:221–235.
  - 33 Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods.* 2002;7:105–125.
  - 34 Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* Thousand Oaks, Calif: Sage; 2004.
  - 35 Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Stat Med.* 2002;21:2131–2144.
  - 36 Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.0.1. Available at: <http://www.cochrane.org/resources/handbook/index.htm>. Accessed May 12, 2010.
  - 37 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557–560.
  - 38 Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ.* 2006;333:597–600.
  - 39 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629–634.
  - 40 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med.* 2003;22:2113–2126.
  - 41 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA.* 2006;295:1152–1160.
  - 42 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
  - 43 Bitzer M. Clinical nursing instruction via the PLATO simulated laboratory. *Nurs Res.* 1966;15:144–150.
  - 44 Cassidy RE, Marshall FJ, Gaston GW, Snodgrass M. Computer assisted instruction for diagnostic problem solving of toothache. *J Dent Educ.* 1972;36:46–56.
  - 45 Mullaney TP, Smith TA, Duell RC, Kaplan A. Four-phase study of computer-assisted and slide-tape methods of stimulating clinical endodontic problems. *J Dent Educ.* 1976;40:681–687.
  - 46 Murray TS, Cupples RW, Barber JH, Dunn WR, Scott DB, Hannay DR. Teaching decision making to medical undergraduates by computer-assisted learning. *Med Educ.* 1977;11:262–264.
  - 47 Schleutermann JA, Holzemer WL, Farrand LL. An evaluation of paper-and-pencil and computer-assisted simulations. *J Nurs Educ.* 1983;22:315–323.
  - 48 Dale RA, Sandoval VA, Hendricson WD, Herbert RJ. A comparison of simulation programs for endodontic review. *J Biocommun.* 1986;12:17–23.
  - 49 Garrett TJ, Ashford AR. Computer-assisted instruction in patient management for internal medicine residents. *J Med Educ.* 1986;61:987–989.
  - 50 Krahn CG, Blanche MC. Using an advance organizer to improve knowledge application by medical students in computer-based clinical simulations. *J Comput Based Instr.* 1986;13:71–74.
  - 51 Sandoval VA, Dale RA, Hendricson WD, Alexander JB. A comparison of four simulation and instructional methods for endodontic review. *J Dent Educ.* 1987;51:532–538.
  - 52 Harless WG, Duncan RC, Zier MA, Ayers WR, Berman JR, Pohl HS. A field test of the TIME patient simulation model. *Acad Med.* 1990;65:327–333.
  - 53 Lyon HC Jr, Healy JC, Bell JR, et al. Computer-based exercises in anemia and chest pain diagnosis: An interim evaluation of the Planalyzer project. *Proc Annu Symp Comput Appl Med Care.* November 7, 1990:483–487.
  - 54 Friedman CP, France CL, Drossman DD. A randomized comparison of alternative formats for clinical simulations. *Med Decis Making.* 1991;11:265–272.
  - 55 Lowdermilk DL, Hopkins Fishel A. Computer simulations as a measure of nursing students' decision-making skills. *J Nurs Educ.* 1991;30:34–39.
  - 56 Lyon HC Jr, Healy JC, Bell JR, et al. Significant efficiency findings while controlling for the frequent confounders of CAI research in the Planalyzer project's computer-based, self-paced, case-based programs in anemia and chest pain diagnosis. *J Med Syst.* 1991;15:117–132.
  - 57 Weverling GJ, Stam J, ten Cate TJ, van Crevel H. Computer-assisted education in problem-solving in neurology: A randomized educational study [in Dutch]. *Ned Tijdschr Geneesk.* 1996;140:440–443.
  - 58 Johnson LA, Cunningham MA, Finkelstein MW, Hand JS. Geriatric patient simulations for dental hygiene. *J Dent Educ.* 1997;61:667–677.
  - 59 Kinney P, Keskula DR, Perry JF. The effect of a computer assisted instructional program on physical therapy students. *J Allied Health.* 1997;26:57–61.
  - 60 Bryce DA, King NJC, Graebner CF, Myers JH. Evaluation of a diagnostic reasoning program (DxR): Exploring student perceptions and addressing faculty concerns. *J Interact Media Educ.* 1998;98:1–34.
  - 61 Schwid HA, Rooke GA, Ross BK, Sivarajan M. Use of a computerized advanced cardiac life support simulator improves retention of advanced cardiac life support guidelines better than a textbook review. *Crit Care Med.* 1999;27:821–824.
  - 62 Fleetwood J, Vaught W, Feldman D, Gracely E, Kassutto Z, Novack D. *MedEthEx online: A computer-based learning program in medical ethics and communications skills.* *Teach Learn Med.* 2000;12:96–104.
  - 63 Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. *Med Educ.* 2001;35:824–832.
  - 64 Schwid HA, Rooke GA, Michalowski P, Ross BK. Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator. *Teach Learn Med.* 2001;13:92–96.
  - 65 Buysse H, Van Maele G, De Moor GJ. The dynamic patient simulator: Learning process, first results and students' satisfaction. *Stud Health Technol Inform.* 2002;93:19–23.
  - 66 Chaikoolvatana A, Goodyer L. Evaluation of a multimedia case-history simulation program for pharmacy students. *Am J Pharm Educ.* 2003;67:108–115.
  - 67 Kumta SM, Tsang PL, Hung LK, Cheng JCY. Fostering critical thinking skills through a Web-based tutorial programme for final year medical students—a randomized, controlled study. *J Educ Multimed Hypermedia.* 2003;12:267–273.
  - 68 Hayes-Roth B, Amano K, Saker R, Sephton T. Training brief intervention with a virtual coach and virtual patients. *Annu Rev Cybertherap Telemed.* 2004;2:85–95.
  - 69 Schitteck Janda M, Mattheos N, Nattestad A, et al. Simulation of patient encounters using a virtual patient in periodontology instruction of dental students: Design, usability, and learning effect in history-taking skills. *Eur J Dent Educ.* 2004;8:111–119.
  - 70 Dickerson R, Johnsen K, Raji A, et al. Virtual patients: Assessment of synthesized versus recorded speech. *Stud Health Technol Inform.* 2006;119:114–119.
  - 71 Ferguson JE 2nd, Kleinert HL, Lunney CA, Campbell LR. Resident physicians' competencies and attitudes in delivering a postnatal diagnosis of Down syndrome. *Obstet Gynecol.* 2006;108:898–905.
  - 72 Raji A, Johnsen K, Dickerson R, et al. Interpersonal scenarios: Virtual ≈ real? Paper presented at: IEEE Virtual Reality Conference; March 2006; Alexandria, Va.
  - 73 Thompson GA, Holyoak KJ, Morrison RG, Clark TK. Evaluation of an online analogical patient simulation program. Paper presented at: IEEE Symposium on

- Computer-Based Medical Systems; June 2006; Salt Lake City, Utah.
- 74 Triola M, Feldman H, Kalet AL, et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. *J Gen Intern Med.* 2006;21:424–429.
  - 75 Turner MK, Simon SR, Facemyer KC, Newhall LM, Veach TL. Web-based learning versus standardized patients for teaching clinical diagnosis: A randomized, controlled, crossover trial. *Teach Learn Med.* 2006;18: 208–214.
  - 76 Wahlgren C-F, Edelbring S, Fors U, Hindbeck H, Stahle M. Evaluation of an interactive case simulation system in dermatology and venereology for medical students. *BMC Med Educ.* 2006;6:40.
  - 77 Deladisma AM, Cohen M, Stevens A, et al. Do medical students respond empathetically to a virtual patient? *Am J Surg.* 2007;193: 756–760.
  - 78 Kleinert HL, Fisher SB, Sanders CL, Boyd S. Improving physician assistant students' competencies in developmental disabilities using virtual patient modules. *J Physician Assist Educ.* 2007;18:33–40.
  - 79 Kleinert HL, Sanders C, Mink J, et al. Improving student dentist competencies and perception of difficulty in delivering care to children with developmental disabilities using a virtual patient module. *J Dent Educ.* 2007;71:279–286.
  - 80 Raji AB, Johnsen K, Dickerson RF, et al. Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Trans Vis Comput Graph.* 2007;13: 443–457.
  - 81 Sanders CL, Kleinert HL, Free T, et al. Caring for children with intellectual and developmental disabilities: Virtual patient instruction improves students' knowledge and comfort level. *J Pediatr Nurs.* 2007;22: 457–466.
  - 82 Sijstermans R, Jaspers MW, Bloemendaal PM, Schoonderwaldt EM. Training inter-physician communication using the dynamic patient simulator. *Int J Med Inform.* 2007;76:336–343.
  - 83 Vash JH, Yunesian M, Shariati M, Keshvari A, Harirchi I. Virtual patients in undergraduate surgery education: A randomized controlled study. *ANZ J Surg.* 2007;77:54–59.
  - 84 Boyd SE, Sanders CL, Kleinert HL, et al. Virtual patient training to improve reproductive health care for women with intellectual disabilities. *J Midwifery Womens Health.* 2008;53:453–460.
  - 85 Sanders C, Kleinert HL, Boyd SE, Herren C, Theiss L, Mink J. Virtual patient instruction for dental students: Can it improve dental care access for persons with special needs? *Spec Care Dentist.* 2008;28:205–213.
  - 86 Sanders CL, Kleinert HL, Free T, King P, Slusher I, Boyd S. Developmental disabilities: Improving competence in care using virtual patients. *J Nurs Educ.* 2008;47: 66–73.
  - 87 Youngblood P, Harter PM, Srivastava S, Moffett S, Heinrichs WL, Dev P. Design, development, and evaluation of an online virtual emergency department for training trauma teams. *Simul Healthc.* 2008;3:146–153.
  - 88 Bearman M. Is virtual the same as real? Medical students' experiences of a virtual patient. *Acad Med.* 2003;78:538–545.
  - 89 Bergin R, Youngblood P, Ayers MK, et al. Interactive simulated patient: Experiences with collaborative e-learning in medicine. *J Educ Comput Res.* 2003;29:387–400.
  - 90 Mallott D, Raczek J, Skinner C, Jarrell K, Shimko M, Jarrell B. A basis for electronic cognitive simulation: The heuristic patient. *Surg Innov.* 2005;12:43–49.
  - 91 de Dombal FT, Hartley JR, Sleeman DH. A computer-assisted system for learning clinical diagnosis. *Lancet.* 1969;1:145–148.
  - 92 Harless WG, Drennon GG, Marxer JJ, Root JA, Miller GE. CASE: A computer-aided simulation of the clinical encounter. *J Med Educ.* 1971;46:443–448.
  - 93 Colliver JA, Kucera K, Verhulst SJ. Meta-analysis of quasi-experimental research: Are systematic narrative reviews indicated? *Med Educ.* 2008;42:858–865.
  - 94 Cook DA. Web-based learning: Pros, cons, and controversies. *Clin Med.* 2007;7:37–42.
  - 95 Oxman A, Guyatt G. When to believe a subgroup analysis. In: Hayward R, ed. *Users' Guides Interactive.* Chicago, Ill: JAMA Publishing Group; 2002.
  - 96 Cook DA, McDonald FS. E-learning: Is there anything special about the "E"? *Perspect Biol Med.* 2008;51:5–21.
  - 97 Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, ed. *The Cambridge Handbook of Multimedia Learning.* New York, NY: Cambridge University Press; 2005:31–48.
  - 98 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;79(10 suppl): S70–S81.
  - 99 Ende J. Feedback in clinical medical education. *JAMA.* 1983;250:777–781.
  - 100 van de Ridder JM, Stokking KM, McGaghie WC, ten Cate OT. What is feedback in clinical education? *Med Educ.* 2008;42:189–197.
  - 101 Alessi SM. Fidelity in the design of instructional simulations. *J Comput Based Instr.* 1988;15:40–47.
  - 102 Cook DA. The research we still are not doing: An agenda for the study of computer-based learning. *Acad Med.* 2005;80:541–548.