

Mastery Learning for Health Professionals Using Technology-Enhanced Simulation: A Systematic Review and Meta-Analysis

David A. Cook, MD, MHPE, Ryan Brydges, PhD, Benjamin Zendejas, MD, MSc, Stanley J. Hamstra, PhD, and Rose Hatala, MD, MSc

Abstract

Purpose

Competency-based education requires individualization of instruction. Mastery learning, an instructional approach requiring learners to achieve a defined proficiency before proceeding to the next instructional objective, offers one approach to individualization. The authors sought to summarize the quantitative outcomes of mastery learning simulation-based medical education (SBME) in comparison with no intervention and nonmastery instruction, and to determine what features of mastery SBME make it effective.

Method

The authors searched MEDLINE, EMBASE, CINAHL, ERIC, PsycINFO,

Scopus, key journals, and previous review bibliographies through May 2011. They included original research in any language evaluating mastery SBME, in comparison with any intervention or no intervention, for practicing and student physicians, nurses, and other health professionals. Working in duplicate, they abstracted information on trainees, instructional design (interactivity, feedback, repetitions, and learning time), study design, and outcomes.

Results

They identified 82 studies evaluating mastery SBME. In comparison with no intervention, mastery SBME was associated with large effects on skills (41 studies; effect size [ES] 1.29 [95%

confidence interval, 1.08–1.50]) and moderate effects on patient outcomes (11 studies; ES 0.73 [95% CI, 0.36–1.10]). In comparison with nonmastery SBME instruction, mastery learning was associated with large benefit in skills (3 studies; effect size 1.17 [95% CI, 0.29–2.05]) but required more time. Pretraining and additional practice improved outcomes but, again, took longer. Studies exploring enhanced feedback and self-regulated learning in the mastery model showed mixed results.

Conclusions

Limited evidence suggests that mastery learning SBME is superior to nonmastery instruction but takes more time.

As clinical medicine grows more complex, medical educators have increasingly recognized the need to explore new paradigms for training,^{1–3}

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

Dr. Brydges is assistant professor, Department of Medicine, University of Toronto, Toronto, Ontario, Canada.

Dr. Zendejas is resident, Department of Surgery, Mayo Clinic College of Medicine, Rochester, Minnesota.

Dr. Hamstra is associate professor, Department of Medicine, Faculty of Medicine, acting assistant dean, Academy for Innovation in Medical Education, and research director, University of Ottawa Skills and Simulation Centre, University of Ottawa, Ottawa, Ontario, Canada.

Dr. Hatala is associate professor, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Correspondence should be addressed to Dr. Cook, Division of General Internal Medicine, Mayo Clinic, 200 First St. SW, Rochester, MN 55905; e-mail: cook.david33@mayo.edu.

Acad Med. 2013;88:00–00.

First published online

doi: 10.1097/ACM.0b013e31829a365d

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A136>.

including tailoring training to individual needs.⁴ Competency-based training⁵ and training milestones⁶ are but two models proposed to achieve that goal at the level of an entire curriculum or educational program. For training to achieve a specific objective (i.e., proficiency in a particular course or task), the mastery learning model offers an analogous and complementary approach. In mastery learning, trainees must achieve a defined proficiency in a given instructional unit before proceeding to the next unit.⁷ Thus, all trainees will meet the same objectives, although learning time typically varies. By contrast, traditional instruction fixes the learning time and allows outcomes to vary.

A recent narrative review of simulation-based medical education (SBME) proposed mastery learning to be an effective instructional design feature⁸ and defined the key characteristics of this model: (1) use of an assessment with an established minimum passing standard, (2) definition of learning objectives aligned with the passing standard, (3) baseline assessment, (4) instruction

that targets learning objectives, (5) reassessment after instruction, (6) progression to the next unit only after achievement of the passing standard, and (7) continued practice if the minimum passing standard was not achieved. However, that review did not offer empiric evidence to support the effectiveness of mastery learning.

In a recent comprehensive meta-analysis of the use of SBME compared with no intervention,⁹ we found a larger effect size for studies using mastery learning than for nonmastery approaches, but these comparisons were indirect (across studies rather than within one study). We are not aware of other quantitative syntheses evaluating mastery learning in health professions education. Looking outside of medical education, a meta-analysis published in 1990 of 108 predominantly college-level studies demonstrated a moderate effect size (0.52) for mastery learning in comparison with nonmastery learning using knowledge outcomes.¹⁰

Given the potential importance of mastery learning in emerging clinical

education models, educators and researchers would benefit from a focused synthesis of evidence for mastery learning in SBME. To address this need, we sought to identify and quantitatively summarize all comparative studies of technology-enhanced simulation using a mastery learning model and involving health professions trainees.

Method

We planned, conducted, and reported this review in adherence with PRISMA standards for reporting meta-analyses.¹¹ We report a planned in-depth analysis of studies from an earlier comprehensive review.⁹ Detailed methods have been reported previously⁹; herein, we abridge these with emphasis on methods unique to the present analyses.

Questions

We sought to answer these questions: What is the effect of mastery learning SBME in comparison with no intervention and nonmastery learning instruction, and what features of mastery learning SBME make it more or less effective? We hypothesized that mastery learning models incorporating instructional design features of cognitive interactivity (promotion of cognitive engagement using strategies such as group discussion or intentional task sequencing),¹² feedback (information on performance provided by an instructor, a peer, or a computer), repetition, and longer time spent learning would be more effective than mastery learning models without these features.

Study eligibility

We defined technology-enhanced simulation as “an educational tool or device with which the learner physically interacts to mimic an aspect of clinical care for the purpose of teaching or assessment.”⁹ This includes mannequins, part-task trainers, virtual reality systems, animal models, and human cadavers but excludes human patient actors (standardized patients) because these are not “technology-enhanced.”

We included quantitative comparative studies published in any language that used a mastery learning model in conjunction with technology-enhanced SBME to teach health professions trainees at any stage in training or practice. We

defined mastery learning as instruction that included all of the key features⁸ listed above. Comparative studies included single-group pretest–posttest studies and studies with one or more control groups or comparison interventions.

Study identification

We previously published our search strategy in full.⁹ To summarize briefly: A research librarian designed a strategy to search MEDLINE, EMBASE, CINAHL, PsycINFO, ERIC, Web of Science, and Scopus for relevant articles. We used no beginning date cutoff, and the last date of search was May 11, 2011. We added the entire reference list from several published reviews of simulation-based education and all articles published in two key journals (*Simulation in Healthcare* and *Clinical Simulation in Nursing*). We sought additional studies from our files and from the reference lists of 190 included articles. We compared our list of included articles against those in a recent review of deliberate practice¹³ and found no omissions.

Study selection

Working independently and in pairs, we screened all articles to identify studies of SBME without regard to instructional approach, reviewing first the titles and abstracts and then the full texts. We resolved conflicts by consensus. As part of data extraction (described below), we subsequently identified all studies that used a mastery model (intraclass correlation, 0.65) for inclusion in the present review.

Data extraction

We worked independently and in pairs, resolving conflicts by consensus, to abstract information on trainee educational level, clinical topic, study design, instructional design (including features of cognitive interactivity, feedback, repetition, and time spent learning), outcomes, and methodological quality. Methodological quality was graded using two previously described instruments, the Medical Education Research Study Quality Instrument¹⁴ (MERSQI) and an adaptation of the Newcastle–Ottawa scale (NOS) for cohort studies.^{15,16}

We abstracted information separately for outcomes of satisfaction, knowledge, skills in an artificial setting, behaviors

with real patients, and effects on patients. We distinguished skill outcomes of time (how long it took to perform a task), process (proficiency during the task, such as global ratings, economy of movements, or minor errors), and product (results observable after the task, such as knot integrity, major complication, or mortality). Time and process behaviors reflected similar measurements in the care of real patients.

Data synthesis

We conducted meta-analyses for mastery SBME versus no intervention and versus non-SBME instruction, and for SBME with versus without a mastery learning model. To ascertain features to guide implementation of mastery learning, we iteratively reviewed included articles to identify salient themes emergent from the literature, conducted meta-analyses for all themes addressed by two or more studies, and performed a critical synthesis of studies comparing two alternative SBME approaches.

For each study outcome, we calculated a standardized mean difference (Hedges g effect size) from the mean and standard deviation (SD), odds ratio, or the results of statistical tests using standard methods as detailed previously.⁹ To facilitate direct comparison with other outcomes, we calculated time effect sizes such that higher numbers indicate favorable results (i.e., less time to complete the task). If articles contained insufficient information to calculate an effect size, we requested this information from authors. To combine the results of all studies making a similar comparison, we pooled effect sizes using random effects. We used the I^2 statistic¹⁷ to quantify how much the results varied across individual studies (i.e., between-study inconsistency, or heterogeneity). To explore anticipated inconsistencies, we planned subgroup analyses for all outcomes with five or more studies evaluating the impact of key instructional design features (cognitive interactivity, feedback, repetitions, and learning time) and key study design features (randomization, assessor blinding, and overall quality score). We performed sensitivity analyses excluding studies that used imputed SDs to estimate the effect size or P value upper limits. We used SAS 9.1 (SAS Institute, Cary, North Carolina) for all analyses. We used funnel plots to explore

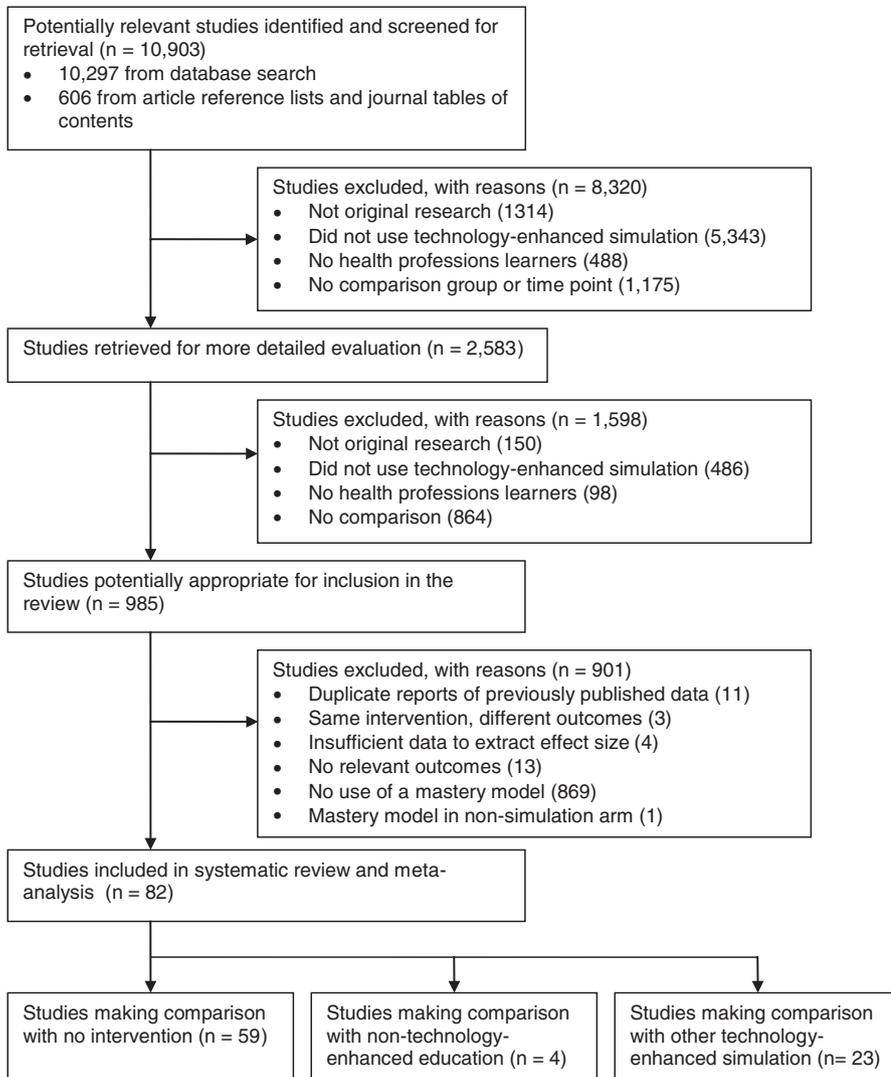


Figure 1 Trial flow in the systematic review of simulation-based mastery learning described in this review.

for possible publication bias, and then in cases of asymmetry, we used trim and fill to calculate revised pooled effect size estimates.* Statistical significance was defined by a two-sided alpha of .05, and interpretations of educational significance emphasized confidence intervals in relation to Cohen effect size classifications (>0.8 = large, 0.5–0.8 = moderate, 0.2–0.5 = small).¹⁸

* Funnel plots are an attempt to evaluate for publication bias by investigating the possibility that small studies showing no statistically significant difference remain unpublished and are therefore omitted from the meta-analysis. An asymmetric funnel plot suggests possible publication bias, but neither the presence nor absence of such bias can truly be known. If publication bias is suspected, trim and fill can be used to estimate the effects of the “missing” (unpublished) studies, and the meta-analysis can be repeated to combine the original data and the new estimates into a revised pooled effect size. However, like the funnel plot, there is no way to verify the accuracy of the trim and fill estimates.

Results

Trial flow

We found 10,903 potentially relevant articles, from which we identified 985 studies of simulation-based health professions education (see Figure 1). Of these, we identified 82 studies employing a mastery learning model in one or more simulation interventions and making comparison with no intervention, with nonsimulation education, and/or with simulation-based education. One additional study used a mastery learning approach only in the nonsimulation intervention, and we do not discuss that study further.¹⁹

Study characteristics

We summarize study characteristics in Table 1, with further details and full citations for all studies in Supplemental

Digital Table 1 and Supplemental List 1; see <http://links.lww.com/ACADMED/A136>. We cite only selected studies in the print narrative of this review.

The 82 studies enrolled a total of 3,498 participants. Forty-nine studies (60%) involved postgraduate trainees, and 26 (32%) involved medical students. The most common clinical topics were minimally invasive surgery, gastrointestinal or urological endoscopy, central or peripheral vascular access, airway management, and resuscitation training. Feedback was high (i.e., provided from multiple sources or with high intensity) in 32 studies (39%), and cognitive interactivity (activities to engage trainees’ thinking) was high in 59 studies (72%). In 38 studies (46%), trainees averaged more than 10 repetitions per task. In 33 studies (40%), instruction lasted five or more hours.

These 82 studies reported 142 discrete outcomes: 4 satisfaction, 4 knowledge, 25 time skills, 63 process skills, 3 product skills, 10 time behaviors (i.e., involving real patients), 18 process behaviors, and 15 patient effects outcomes. Skills were usually assessed with the same simulator as was used in training (e.g., using the same laparoscopic surgery simulator for both training and assessment). However, 9 of 25 (36%) time skills measures, 20 of 63 (32%) process skills measures, and 2 of 3 (67%) product skills measures assessed skills using another simulation modality (e.g., training with a laparoscopic surgery box trainer and assessing with a live pig).

Study quality

We summarize study quality in Table 2 and Supplemental Digital Table 1 (which can be accessed at <http://links.lww.com/ACADMED/A136>). Twenty-five studies (31%) used a single-group pre–post design. Of the studies with a comparison arm, 42 (74%) used random group allocation. Twenty-six (32%) studies reported >25% attrition after enrollment or failed to report attrition. We found evidence to support the validity of assessment scores in terms of content, internal structure, and relations with other variables in approximately one-third of the reports (see Table 2). Seventy-eight of the 142 outcome measures (55%) were blinded. Sixteen outcome measures (11%) were reported by the trainee; the rest were determined

Table 1

Description of Studies Included in a Systematic Review of Simulation-Based Mastery Learning*

Study characteristic	Level	No. of studies (no. of participants [†]), by comparison [‡]			
		Any comparison	No intervention	Nonsimulation	Other simulation
All studies		82 (3,498)	59 (2,214)	4 (395)	23 (1,050)
Study design	Posttest-only 2-group	33 (1,814)	22 (959)	3 (360)	10 (564)
	Pretest-posttest 2-group	24 (818)	12 (389)	1 (35)	13 (486)
	Pretest-posttest 1-group	25 (866)	25 (866)	0	0
Group allocation	Randomized	42 (1,538)	23 (730)	3 (95)	20 (874)
Participants [‡]	Medical students	26 (833)	15 (528)	2 (71)	12 (302)
	Physicians postgraduate training	49 (1,463)	42 (1,263)	1 (24)	7 (216)
	Physicians in practice	7 (59)	6 (59)	0	2 (20)
	Nurses and nursing students	4 (232)	3 (172)	0	1 (60)
	Emergency medical technicians and students	4 (247)	1 (missing [¶])	0	3 (247)
	Dentists and dental students	2 (94)	1 (78)	0	1 (16)
	Other / ambiguous / mixed	7 (570)	3 (114)	1 (300)	4 (189)
	Clinical topics [§]	Minimally invasive surgery	40 (974)	29 (692)	1 (35)
	Endoscopy or ureteroscopy	9 (171)	8 (135)	0	1 (36)
	Vascular access	8 (556)	6 (476)	0	2 (80)
	Airway management	7 (522)	2 (66)	1 (24)	4 (432)
	Resuscitation / trauma training	6 (359)	4 (257)	0	2 (102)
	Other surgery	6 (354)	5 (318)	1 (36)	1 (36)
	Physical examination	2 (312)	2 (312)	0	0
Outcomes [‡]	Satisfaction	4 (407)	0	1 (285)	3 (122)
	Knowledge	4 (486)	2 (148)	2 (315)	1 (53)
	Skill: time	25 (498)	15 (278)	3 (81)	8 (193)
	Skill: process	63 (2,116)	41 (1,523)	4 (111)	20 (624)
	Skill: product	3 (80)	3 (80)	0	0
	Behavior: time	10 (178)	9 (154)	1 (24)	0
	Behavior: process	18 (875)	14 (533)	3 (339)	2 (33)
	Patient effects	15 (1,111)	11 (537)	2 (309)	2 (154)
Quality	Newcastle–Ottawa ≥ 4 points	33 (1,590)	17 (590)	4 (395)	14 (673)
	MERSQI ≥ 12 points	62 (2,698)	43 (1,611)	4 (395)	18 (793)

* The authors conducted a systematic review of studies evaluating the outcomes of mastery learning simulation-based medical education.

[†] Numbers reflect the number enrolled, except for the numbers in the Outcomes rows, which reflect number of participants who provided observations for analysis.

[‡] The number of studies and trainees in some subgroups (summing across rows or columns) may sum to more than the number for all studies because several studies included more than one comparison arm or more than one trainee group, fit within more than one clinical topic, or reported multiple outcomes.

[§] Selected listing of the topics addressed most often (numerous other topics were addressed, with lower frequency). Some interventions addressed more than one clinical topic.

[¶] The number of participants in this study was reported only for entire sample, not separated by role; participants counted under "other/ambiguous/mixed."

by computer or instructor ratings. The mean (SD) MERSQI score was 12.8 (2.0) on an 18-point scale; the mean NOS score was 3.0 (1.6) on a 6-point scale.

Quantitative synthesis

Meta-analysis: SBME mastery learning versus no intervention. Fifty-nine studies (enrolling 2,214 trainees) compared mastery learning SBME with no

intervention. All but 2 of the 95 outcomes showed benefit. Both exceptions involved time outcomes, and in both studies other outcomes showed benefit (i.e., performance improved but took longer).

Figure 2 shows the pooled effect size for each outcome. For the most prevalent outcome, process skills (41 studies, 1,523 trainees), we found a pooled effect size of 1.29 (95% confidence interval [CI],

1.08–1.50; $P < .001$). Because effect sizes greater than 0.8 are considered large,¹⁸ this suggests that mastery learning SBME is associated with substantial learning gains compared with no intervention. However, we also found large inconsistency among studies ($I^2 = 81\%$), with individual effect sizes ranging from 0.22 to 4.56. We explored this inconsistency by performing subgroup analyses to determine whether high cognitive interactivity, high feedback,

Table 2
Quality of Studies Included in a Systematic Review of Simulation-Based Mastery Learning*

Scale item	Subscale (points if present)	No. (%) present; N = 82
Medical Education Research Study Quality Instrument (MERSQI)[†]		
Study design (maximum 3)	1-group pre-post (1.5)	25 (31)
	Observational 2-group (2)	15 (18)
	Randomized 2-group (3)	42 (51)
Sampling: no. of institutions (maximum 1.5)	1 (0.5)	68 (83)
	2 (1)	4 (5)
	>2 (1.5)	10 (12)
Sampling: follow-up (maximum 1.5)	<50% or not reported (0.5)	23 (28)
	50%–74% (1)	3 (4)
	≥75% (1.5)	56 (68)
Type of data: outcome assessment (maximum 3)	Subjective (1)	7 (9)
	Objective (3)	75 (91)
Validity evidence (maximum 3)	Content (1)	32 (39)
	Internal structure (1)	29 (35)
	Relations to other variables (1)	29 (35)
Data analysis: appropriate (maximum 1)	Appropriate (1)	74 (90)
Data analysis: sophistication (maximum 2)	Descriptive (1)	3 (4)
	Beyond descriptive analysis (2)	79 (96)
Highest outcome type (maximum 3)	Knowledge, skills (1.5)	58 (71)
	Behaviors (2)	9 (11)
	Patient/health care outcomes (3)	15 (18)
Newcastle–Ottawa Scale (modified)[‡]		
Representativeness of sample	Present (1)	20 (24)
Comparison group from same community	Present (1)	52 (63)
Comparability of comparison cohort, criterion A [§]	Present (1)	43 (52)
Comparability of comparison cohort, criterion B [§]	Present (1)	24 (29)
Blinded outcome assessment	Present (1)	47 (57)
Follow-up high	Present (1)	57 (70)

*The authors conducted a systematic review of studies evaluating the outcomes of mastery learning simulation-based medical education.

[†] Mean (SD) MERSQI score was 12.8 (2.0); median (range) was 12.75 (6–17).

[‡] Mean (SD) Newcastle–Ottawa Scale score was 3.0 (1.6); median (range) was 3 (0–6).

[§] Comparability of cohorts criterion A was present if the study (1) was randomized or (2) controlled for a baseline learning outcome; criterion B was present if (1) a randomized study concealed allocation or (2) an observational study controlled for another baseline trainee characteristic.

multiple repetitions (>10 versus ≤9), or more time spent learning (≥5 hours versus <5) influenced process skills outcomes. (See full results in Supplemental Digital Table 2, which can be accessed at <http://links.lww.com/ACADMED/A136>.) In each case, the interaction was not statistically significant, suggesting that these instructional features were not associated with the outcomes studied. We also explored the impact of study methods in analyses grouped by randomization, blinding, and high/low MERSQI and NOS scores. These subgroups were not statistically significant except that studies with high NOS scores (≥4) had

lower outcomes than lower-quality studies (pooled effect size 0.97 versus 1.40, $P_{interaction} = .049$). Sensitivity analyses excluding two studies with imprecise effect sizes yielded results virtually identical to those of the main analysis. A visibly asymmetric funnel plot suggested possible publication bias. Assuming this asymmetry does reflect publication bias, trim and fill analyses yielded a slightly lower but still large effect size (1.14).

Eleven studies (537 trainees) reported outcomes reflecting direct impact on patients such as procedural success,

patient satisfaction, and complications. For these outcomes, mastery learning SBME was associated with a moderate pooled effect size of 0.73 (95% CI, 0.36–1.10; $P < .001$). Inconsistency was large ($I^2 = 55\%$), and effect sizes ranged from 0.09 to 1.68. In subgroup analyses, high cognitive interactivity was associated with higher outcomes (pooled effect size 0.88 versus 0.16, $P_{interaction} = .015$). No other subgroup analysis interactions were statistically significant (see Supplemental Digital Table 2, which can be accessed at <http://links.lww.com/ACADMED/A136>). The funnel plot was symmetric.

We found large pooled effect sizes for all other outcomes except product skills, which showed a moderate effect (see Figure 2). Inconsistency was also large in nearly all analyses. Subgroup analyses did not reveal a consistent pattern of effect (see Supplemental Digital Table 2, <http://links.lww.com/ACADMED/A136>). The pooled effect sizes for time skills were significantly higher for studies with nonrandomized allocation, low NOS scores, and few repetitions. For time behaviors, pooled effect sizes were significantly higher for low cognitive interactivity and high NOS scores. For process behaviors, pooled effect sizes were significantly higher for long learning times and blinded outcome measures. The funnel plots for time skills, time behaviors, and process behaviors were visibly asymmetric. Again assuming this asymmetry reflects publication bias, we found revised effect sizes moderate in magnitude for time skills (0.69) and time behaviors (0.65) and large for process behaviors (0.82).

Meta-analysis: SBME mastery learning versus nonsimulation instruction. Four studies compared mastery learning SBME with nonsimulation instruction (lecture or video).^{20–23} Individual or pooled effect sizes for these studies (see Figure 3) were moderate to large in favor of SBME.

Meta-analysis: SBME with and without a mastery learning model. Five studies directly compared SBME using a mastery model with nonmastery SBME (see Figure 4).^{24–28} For the three studies reporting a process skills outcome, results favor the mastery model with a pooled effect size of 1.17 (95% CI, 0.29–2.05; $P = .009$) and $I^2 = 74\%$.

Two of those five studies reported direct patient effects, and one reported patient-

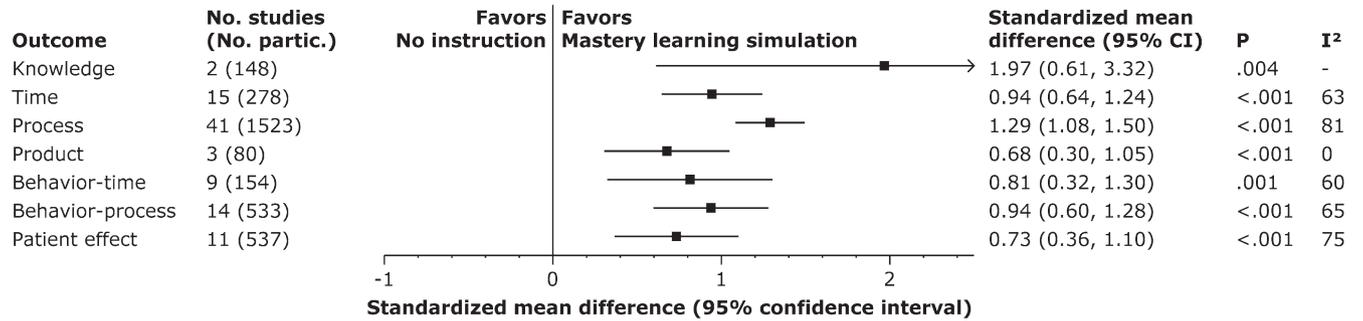


Figure 2 Outcomes of studies comparing simulation-based mastery learning versus no intervention. Positive standardized mean differences favor the simulation intervention. Results reflect pooling using random-effects meta-analysis. Time outcomes refer to performance (not training time); positive numbers indicate favorable effects (faster performance).

related behaviors. The pooled results of these three studies again favor the mastery learning model, although the effect is not statistically significant, with an effect size of 0.26 (95% CI, -0.07 to 0.58; $P = .12$) and low inconsistency ($I^2 = 0\%$).

Two studies reported information on the duration of instruction.^{27,28} As would be expected, instruction with the mastery learning model took longer, requiring more time in one study²⁸ (51 versus 48 minutes [effect size 0.25]) and more repetitions in the other²⁷ (62 versus 42 [effect size 0.55]).

Meta-analysis: SBME mastery learning with and without additional practice. We identified one emergent theme addressed by multiple studies: the addition of extra practice (pretraining,²⁹⁻³¹ additional repetitions,^{32,33} or mental rehearsal³⁴) to mastery learning SBME common in both interventions. Across these six studies, the pooled effect size for process skills outcomes was 0.52 favoring additional practice (95% CI, -0.05 to 1.09; $P = .076$), and inconsistency was high ($I^2 = 83\%$).

For the three studies reporting time skills, the pooled effect size was 1.03 (95% CI, 0.47 to 1.59, $P = .0003$), and inconsistency was low ($I^2 = 13\%$).

Three of these studies reported information on the duration of instruction.²⁹⁻³¹ As would be expected, total learning time was greater for mastery learning with (versus without) additional pretraining or repetitions (87 versus 61 minutes [ES 1.45], 10.4 versus 9.2 hours [ES 0.43], and 351 versus 310 minutes [ES 0.49]).

Cost of instruction. Four studies reported the cost of mastery learning SBME in comparison with another intervention.^{20,30,31,35} A national training program for intrauterine device insertion found improved learning outcomes using a mastery learning SBME model, while saving money because of a shorter overall training time.²⁰ A study found essentially identical learning outcomes but substantial cost savings by using an inexpensive physical pelvic model rather than a virtual reality system.³⁵ Two studies found overall cost savings when learners pretrained on a less expensive simulator,

with generally favorable impact on outcomes.^{30,31}

Critical synthesis: Exploring additional features of mastery learning

Self-regulation. Self-regulated learning allows trainees to monitor and respond to their own instructional needs.³⁶ The circumstances in which health professions trainees can effectively self-regulate, and when they need input from external sources (e.g., human preceptors), are not yet fully understood. Three randomized trials evaluated trainees' ability to self-regulate the mastery learning experience.^{27,28,37}

Two of these studies^{27,28} evaluated trainees' capacity to self-determine mastery rather than rely on an external measure, and found conflicting results. One of these studies²⁸ found that trainees in a one-day training session who determined for themselves when to advance to the next task performed similarly to trainees required to demonstrate objective mastery before advancing. By contrast, the other study²⁷ found that trainees with defined proficiency targets performed better after four months of

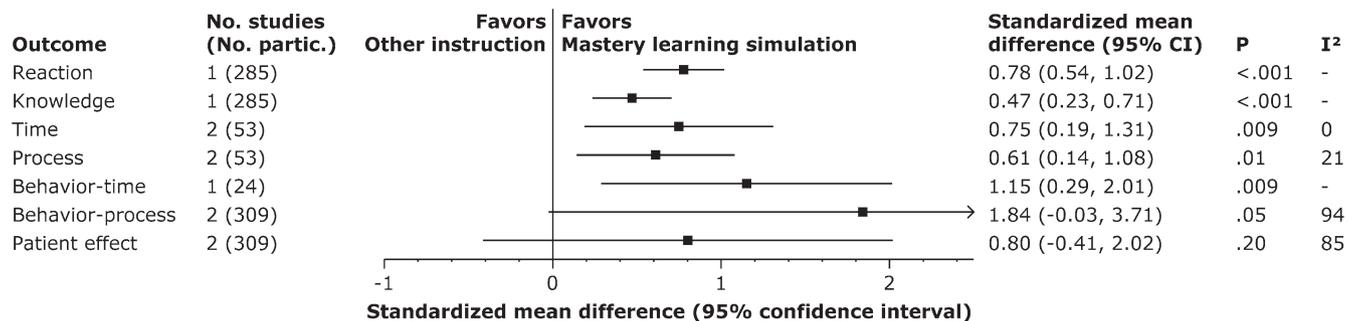


Figure 3 Outcomes of studies comparing simulation-based mastery learning versus other instruction (four studies in all, some reporting multiple outcomes); positive standardized mean differences favor the simulation intervention. Results reflect meta-analytic pooling when more than one study. Time outcomes refer to performance (not training time); positive numbers indicate favorable effects (faster performance).

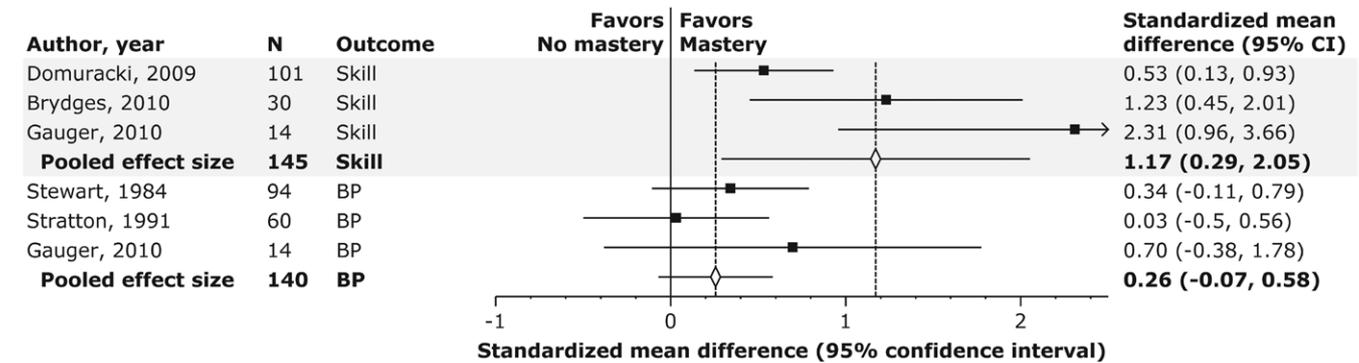


Figure 4 Meta-analysis of studies of simulation-based learning with and without a mastery learning model. Positive standardized mean differences favor simulation with the mastery model. Two separate meta-analyses are reflected in this figure: Skill indicates process skills; $I^2 = 74\%$ for this analysis; pooled effect size $P = .009$. BP indicates behaviors and patient effects; $I^2 = 0\%$ for this analysis; pooled effect size $P = .12$.

training than did those without such targets. The timing of instruction may account for these divergent results, that is, an external standard may be more important for longitudinal training than for short courses.

In the third study,³⁷ trainees who received feedback from a virtual reality simulator performed similarly to those who received feedback from both the simulator and a human preceptor. This suggests that in some circumstances trainees can implement simulator feedback without assistance from a human preceptor.

Feedback. Feedback plays an essential role in learning³⁸ and is critical to the mastery learning model. Four studies explored varying approaches to feedback in conjunction with mastery learning.^{23,26,37,39} Three of these explored enhanced or continuous feedback. One study²³ randomized trainees to practice suturing using a pig's foot (low feedback) or using an investigator-developed model that provided visual (colored lights) and audible (tone) feedback. Those using the new model (enhanced feedback) had better and faster performance following training. In a randomized trial of cricoid pressure instruction,²⁶ one group trained under a mastery learning model while receiving continuous feedback, while the other group trained with minimal feedback and no mastery requirement. When tested, the continuous feedback group had superior performance. By contrast, a nonrandomized study³⁹ found that continuous feedback during laparoscopy training was inferior to feedback limited to 10 minutes per session. This paradoxical finding could be explained by the guidance hypothesis,⁴⁰ which proposes that trainees can become

dependent on continuous feedback, and performance then suffers when feedback is withdrawn.

The fourth study was mentioned above in relation to self-regulation³⁷ and found that simulator-based feedback was similar in effectiveness to combined simulator and human feedback.

Additional instructional design features. Six studies compared different simulation modalities in the context of mastery learning.^{23,35,41–44} One study each contrasted residents or staff physicians as teachers,⁴⁵ training under high- or low-stress conditions,⁴⁶ training tasks of high or low clinical relevance,⁴⁷ and training with only the dominant hand or with both the dominant and nondominant hand.⁴⁸ Given the diversity of themes and results, we will not discuss these herein.

Discussion

In studies directly comparing SBME with and without a mastery model, mastery learning was associated with higher outcomes. For process skills outcomes, the effect was large and statistically significant. For behaviors and patient effects, the effect was small and not statistically significant, but consistently positive in all three studies. However, as would be expected, mastery learning took longer than nonmastery learning.

In comparison with no intervention, mastery learning SBME was consistently associated with better learning outcomes. Pooled effect sizes were moderate for product skills and patient effect outcomes, and large for all other outcomes, including patient-related behaviors. Subgroup analyses exploring high inconsistency

confirmed effects of similar size across most of the design variations tested. Mastery learning SBME was also more effective than non-SBME instruction.

The included studies provide only scant evidence to clarify our understanding of how to optimally use mastery SBME. Limited evidence supports the use of pretraining, enhanced feedback, and clinically relevant tasks, although continuous feedback may paradoxically impede learning. Evidence regarding self-regulation is too preliminary to permit conclusions.

Limitations and strengths

The studies in this review shared the common theme of mastery learning with SBME, but as has been found in other meta-analytic syntheses of education research,^{9,16,49} between-study variability (inconsistency) was large in most analyses. This likely results from between-study differences in trainee level, simulation modality, clinical topic, outcome measure, and other features of instructional design. This diversity is a weakness in terms of between-study inconsistency, but a strength in terms of comprehensiveness and breadth of scope. Moreover, despite the quantitatively high inconsistency, the comparisons with no intervention favored mastery SBME for all but two outcomes, and comparisons with nonmastery SBME consistently favored mastery, indicating that studies varied in the magnitude but not the direction of benefit.

As in any review, our results are limited by the quantity and quality of the original studies. We found many studies making a comparison with no intervention,

but few making a comparison with an active intervention. This paucity of evidence limits the strength of the inferences regarding mastery learning in comparison with nonmastery learning.

Subgroup analyses should be interpreted with caution⁵⁰ because such between-study comparisons are blurred by simultaneous variation in other study features such as learners, topics, and outcome measures. Moreover, we found inconsistent results across outcomes. The evaluation and adjustments for publication bias are likewise limited in the presence of inconsistency.^{51,52}

Although all of the studies in this review met established criteria for mastery learning,⁸ we nonetheless found diversity in the specific implementation of this model. We made no attempt, however, to judge the appropriateness of the mastery criterion or standardization of the test.

Our review has several strengths, including a comprehensive literature search, rigorous and reproducible coding, and focused analyses. By using broad initial inclusion criteria (i.e., focused on SBME generally) and then identifying mastery learning interventions within this large pool of articles, we identified many relevant studies that would have been missed using a search with more focused terms.

Comparison with previous reviews

Issenberg and colleagues⁵³ seminal review proposed individualized instruction as a key feature of effective simulation on the basis of its prevalence in the literature. Our quantitative synthesis confirms the benefits of individualization in the form of mastery learning. A previous meta-analysis of SBME with deliberate practice (a learning model related to the mastery model, focusing on the instructional phase) found, similar to our review, large favorable effects for 14 studies comparing SBME with no intervention.¹³ All of these deliberate practice studies met our definition of mastery learning and were included in this review.

Our results agree in general with those of a previous review of simulation-based education,⁹ and other reviews of educational technologies in medical education,^{16,49} in that studies making comparison with no intervention typically

find large effects. We maintain that no-intervention-comparison studies do little to advance the science of education, and we suggest that researchers focus on questions that clarify when and how to use these educational technologies.⁵⁴

Implications

Our findings have important implications for current practice and future research. Mastery learning SBME is effective, and educators should consider using this approach as appropriate. The mastery model may be particularly relevant to competency-based education, given the shared emphasis on defined objectives rather than defined learning time. Indeed, these complementary models differ primarily in their focus (a single instructional unit versus an entire training program). However, the optimal role for mastery learning has yet to be defined. For example, evidence supporting the use of mastery SBME for nonprocedural clinical topics is currently lacking.

Much remains to be learned about the optimal implementation of mastery learning SBME. Each of the key tenets of the mastery model⁸ would benefit from further clarification—for example, how to develop the passing standard, instructional objectives, and assessment tools; how to implement the practice phase and (when needed) continued practice; and how to regulate advancement to the next unit. As summarized above, few studies in medical education have directly investigated these issues, and between-study comparisons (e.g., subgroup analyses) are inefficient. Because differences in instructional design and assessment have potentially substantial implications for learning efficiency and optimal use of faculty time, head-to-head comparative effectiveness studies evaluating such features will be essential.

Although mastery learning improves outcomes, it comes at the price of increased learning time and may impose additional logistic burdens on teachers and trainees. Curriculum planners will find it impossible to incorporate all desirable educational activities (including mastery learning) given finite time and resources. Educators must thus consider the efficiencies and comparative value of potential training activities, including

both the benefits of training and the costs in terms of time (of trainees, instructors, and other personnel), money, and lost opportunities (other worthwhile activities that could have been pursued). Few studies to date have evaluated the costs of mastery learning SBME, and this warrants greater attention in future research as we try to understand and maximize the true value of this instructional approach.

Acknowledgments: The authors thank Jason H. Szostek, MD, Amy T. Wang, MD, and Patricia J. Erwin, MLS, for their assistance in the literature search and initial data acquisition.

Funding/Support: This work was supported by intramural funds, including an award from the Division of General Internal Medicine, Mayo Clinic.

Other disclosures: None.

Ethical approval: Not applicable, as no human subjects were involved.

References

Note: For references 55–110, which were not included in the text of this report, see Supplemental Digital List 1. That list can be found at <http://links.lww.com/ACADMED/A136>.

- 1 Emanuel EJ, Fuchs VR. Shortening medical training by 30%. *JAMA*. 2012;307:1143–1144.
- 2 Hodges BD. A tea-steeping or i-Doc model for medical education? *Acad Med*. 2010;85 (9 suppl):S34–S44.
- 3 Ludmerer KM, Johns MM. Reforming graduate medical education. *JAMA*. 2005;294:1083–1087.
- 4 Cooke M, Irby DM, O'Brien BC. *Educating Physicians: A Call for Reform of Medical School and Residency*. San Francisco, Calif: Jossey-Bass; 2010.
- 5 Weinberger SE, Pereira AG, Iobst WF, Mehaber AJ, Bronze MS; Alliance for Academic Internal Medicine Education Redesign Task Force II. Competency-based education and training in internal medicine. *Ann Intern Med*. 2010;153:751–756.
- 6 Green ML, Aagaard EM, Caverzagie KJ, et al. Charting the road to competence: Developmental milestones for internal medicine residency training. *J Grad Med Educ*. 2009;1:5–20.
- 7 Block JH, Burns RB. Mastery learning. *Rev Res Educ*. 1976;4:3–49.
- 8 McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. *Med Educ*. 2010;44:50–63.
- 9 Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA*. 2011;306:978–988.
- 10 Kulik C-LC, Kulik JA, Bangert-Drowns RL. Effectiveness of mastery learning programs: A meta-analysis. *Rev Educ Res*. 1990;60: 265–299.
- 11 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, W64.
- 12 Cook DA, Hamstra SJ, Brydges R, et al. Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Med Teach.* 2013;35:e867–e898.
 - 13 McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med.* 2011;86:706–711.
 - 14 Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA.* 2007;298:1002–1009.
 - 15 Wells GA, Shea B, O'Connell D, et al. The Newcastle–Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm. Accessed April 18, 2013.
 - 16 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.
 - 17 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557–560.
 - 18 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
 - 19 Bonnetain E, Boucheix JM, Hamet M, Freysz M. Benefits of computer screen-based simulation in learning cardiac arrest procedures. *Med Educ.* 2010;44:716–722.
 - 20 Limpaphayom K, Ajello C, Reinprayoon D, Lumbiganon P, Graffkin L. The effectiveness of model-based training in accelerating IUD skill acquisition. A study of midwives in Thailand. *Br J Fam Plann.* 1997;23:58–61.
 - 21 Naik VN, Matsumoto ED, Houston PL, et al. Fiberoptic orotracheal intubation on anesthetized patients: Do manipulation skills learned on a simple model transfer into the operating room? *Anesthesiology.* 2001;95:343–348.
 - 22 Tanoue K, Yasunaga T, Konishi K, et al. Effectiveness of training for endoscopic surgery using a simulator with virtual reality: Randomized study. *Int Congr Ser.* 2005;1281:515–520.
 - 23 Salvendy G, Pilitsis J. The development and validation of an analytical training program for medical suturing. *Hum Factors.* 1980;22:153–170.
 - 24 Stewart RD, Paris PM, Pelton GH, Garretson D. Effect of varied training techniques on field endotracheal intubation success rates. *Ann Emerg Med.* 1984;13:1032–1036.
 - 25 Stratton SJ, Kane G, Gunter CS, et al. Prospective study of manikin-only versus manikin and human subject endotracheal intubation training of paramedics. *Ann Emerg Med.* 1991;20:1314–1318.
 - 26 Domuracki KJ, Moule CJ, Owen H, Kostandoff G, Plummer JL. Learning on a simulator does transfer to clinical practice. *Resuscitation.* 2009;80:346–349.
 - 27 Gauger PG, Hauge LS, Andreatta PB, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. *Am J Surg.* 2010;199:72–80.
 - 28 Brydges R, Carnahan H, Rose D, Dubrowski A. Comparing self-guided learning and educator-guided learning formats for simulation-based clinical training. *J Adv Nurs.* 2010;66:1832–1844.
 - 29 Lammers RL. Learning and retention rates after training in posterior epistaxis management. *Acad Emerg Med.* 2008;15:1181–1189.
 - 30 Rosenthal ME, Castellvi AO, Goova MT, Hollett LA, Dale J, Scott DJ. Pretraining on Southwestern stations decreases training time and cost for proficiency-based fundamentals of laparoscopic surgery training. *J Am Coll Surg.* 2009;209:626–631.
 - 31 Stefanidis D, Hope WW, Korndorffer JR Jr, Markley S, Scott DJ. Initial laparoscopic basic skills training shortens the learning curve of laparoscopic suturing and is cost-effective. *J Am Coll Surg.* 2010;210:436–440.
 - 32 Kovacs G, Bullock G, Ackroyd-Stolarz S, Cain E, Petrie D. A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Ann Emerg Med.* 2000;36:301–309.
 - 33 Stefanidis D, Korndorffer JR Jr, Markley S, Sierra R, Scott DJ. Proficiency maintenance: Impact of ongoing simulator training on laparoscopic skill retention. *J Am Coll Surg.* 2006;202:599–603.
 - 34 Arora S, Aggarwal R, Sirimanna P, et al. Mental practice enhances surgical technical skills: A randomized controlled study. *Ann Surg.* 2011;253:265–270.
 - 35 McDougall EM, Kolla SB, Santos RT, et al. Preliminary study of virtual reality and model simulation for learning laparoscopic suturing skills. *J Urol.* 2009;182:1018–1025.
 - 36 Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. *Med Educ.* 2012;46:71–79.
 - 37 Snyder CW, Vandromme MJ, Tyra SL, Porterfield JR Jr, Clements RH, Hawn MT. Effects of virtual reality simulator training method and observational learning on surgical performance. *World J Surg.* 2011;35:245–252.
 - 38 van de Ridder JM, Stokking KM, McGaghie WC, ten Cate OT. What is feedback in clinical education? *Med Educ.* 2008;42:189–197.
 - 39 Stefanidis D, Korndorffer JR Jr, Heniford BT, Scott DJ. Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. *Surgery.* 2007;142:202–206.
 - 40 Lee TD, White MA, Carnahan H. On the role of knowledge of results in motor learning: Exploring the guidance hypothesis. *J Mot Behav.* 1990;22:191–208.
 - 41 Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surg Endosc.* 2002;16:406–411.
 - 42 Hoadley TA. Learning advanced cardiac life support: A comparison study of the effects of low- and high-fidelity simulation. *Nurs Educ Perspect.* 2009;30:91–95.
 - 43 Scerbo MW, Bliss JP, Schmidt EA, Thompson SN. The efficacy of a medical virtual reality simulator for training phlebotomy. *Hum Factors.* 2006;48:72–84.
 - 44 Thompson JR, Leonard AC, Doarn CR, Roesch MJ, Broderick TJ. Limited value of haptics in virtual reality laparoscopic cholecystectomy training. *Surg Endosc.* 2011;25:1107–1114.
 - 45 Rosenthal ME, Adachi M, Ribaldo V, Mueck JT, Schneider RF, Mayo PH. Achieving housestaff competence in emergency airway management using scenario based simulation training: Comparison of attending vs housestaff trainers. *Chest.* 2006;129:1453–1458.
 - 46 Stefanidis D, Korndorffer JR Jr, Markley S, Sierra R, Heniford BT, Scott DJ. Closing the gap in operative performance between novices and experts: Does harder mean better for laparoscopic simulator training? *J Am Coll Surg.* 2007;205:307–313.
 - 47 Uchal M, Raftopoulos Y, Tjugum J, Bergamaschi R. Validation of a six-task simulation model in minimally invasive surgery. *Surg Endosc.* 2005;19:109–116.
 - 48 Molinas CR, Campo R. Defining a structured training program for acquiring basic and advanced laparoscopic psychomotor skills in a simulator. *Gynecol Surg.* 2010;7:427–435.
 - 49 Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Acad Med.* 2010;85:1589–1602.
 - 50 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med.* 1992;116:78–84.
 - 51 Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ.* 2006;333:597–600.
 - 52 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med.* 2003;22:2113–2126.
 - 53 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.
 - 54 Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: A framework for classifying the purposes of research in medical education. *Med Educ.* 2008;42:128–133.