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 associate 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/VACADMED/A136.
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. We report a planned in-depth analysis of studies from an earlier comprehensive review. Detailed methods have been reported previously; 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), 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." 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 features 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. 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 of 190 included articles. We abstracted information separately to the present reviews.

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 Instrument and an adaptation of the Newcastle–Ottawa scale (NOS) for cohort studies.

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. 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 I² statistic 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...
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 by the trainee; the rest were determined
Review

Academic Medicine, Vol. 88, No. 08 / August 2013

4

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 (I² = 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,
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* = .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 of 0.88 versus 1.06, *P* = .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 3). 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 I² = 74%.

Two of those five studies reported direct patient effects, and one reported patient-
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 (I² = 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 (I² = 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 (I² = 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.26 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 studies27,28 evaluated trainees’ capacity to self-determine mastery rather than rely on an external measure, and found conflicting results. One of these studies 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 study found that trainees with defined proficiency targets performed better after four months of

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).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. studies (No. partic.)</th>
<th>Favors</th>
<th>Favors</th>
<th>Standardized mean difference (95% CI)</th>
<th>P</th>
<th>I²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>2 (148)</td>
<td>No instruction</td>
<td>Mastery learning simulation</td>
<td>1.97 (0.61, 3.32)</td>
<td>.004</td>
<td>-</td>
</tr>
<tr>
<td>Time</td>
<td>15 (278)</td>
<td></td>
<td></td>
<td>0.94 (0.64, 1.24)</td>
<td>&lt;.001</td>
<td>63</td>
</tr>
<tr>
<td>Process</td>
<td>41 (1523)</td>
<td></td>
<td></td>
<td>1.29 (1.08, 1.50)</td>
<td>&lt;.001</td>
<td>81</td>
</tr>
<tr>
<td>Product</td>
<td>3 (80)</td>
<td></td>
<td></td>
<td>0.68 (0.30, 1.05)</td>
<td>&lt;.001</td>
<td>0</td>
</tr>
<tr>
<td>Behavior-time</td>
<td>9 (154)</td>
<td></td>
<td></td>
<td>0.81 (0.32, 1.30)</td>
<td>.001</td>
<td>60</td>
</tr>
<tr>
<td>Behavior-process</td>
<td>14 (533)</td>
<td></td>
<td></td>
<td>0.94 (0.60, 1.28)</td>
<td>&lt;.001</td>
<td>65</td>
</tr>
<tr>
<td>Patient effect</td>
<td>11 (537)</td>
<td></td>
<td></td>
<td>0.73 (0.36, 1.10)</td>
<td>&lt;.001</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. studies (No. partic.)</th>
<th>Favors</th>
<th>Favors</th>
<th>Standardized mean difference (95% CI)</th>
<th>P</th>
<th>I²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction</td>
<td>1 (285)</td>
<td></td>
<td></td>
<td>0.78 (0.54, 1.02)</td>
<td>&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td>Knowledge</td>
<td>1 (285)</td>
<td></td>
<td></td>
<td>0.47 (0.23, 0.71)</td>
<td>&lt;.001</td>
<td>-</td>
</tr>
<tr>
<td>Time</td>
<td>2 (53)</td>
<td></td>
<td></td>
<td>0.75 (0.19, 1.31)</td>
<td>.009</td>
<td>0</td>
</tr>
<tr>
<td>Process</td>
<td>2 (53)</td>
<td></td>
<td></td>
<td>0.61 (0.14, 1.08)</td>
<td>.01</td>
<td>21</td>
</tr>
<tr>
<td>Behavior-time</td>
<td>1 (24)</td>
<td></td>
<td></td>
<td>1.15 (0.29, 2.01)</td>
<td>.009</td>
<td>-</td>
</tr>
<tr>
<td>Behavior-process</td>
<td>2 (309)</td>
<td></td>
<td></td>
<td>1.84 (0.93, 3.71)</td>
<td>.05</td>
<td>94</td>
</tr>
<tr>
<td>Patient effect</td>
<td>2 (309)</td>
<td></td>
<td></td>
<td>0.80 (-0.41, 2.02)</td>
<td>.20</td>
<td>85</td>
</tr>
</tbody>
</table>

Figure 3  Outcomes of studies comparing simulation-based mastery learning versus other instruction. Simulation compared with other (nonsimulation) 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).
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 learning48 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,39 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-regulation35 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,4,5,6,10 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 caution 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.

Although all of the studies in this review met established criteria for mastery learning, 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 colleagues' 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. 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, and other reviews of educational technologies in medical education, 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.

**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 model 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.

11. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: