Adv in Health Sci Educ
DOI 10.1007/s10459-013-9462-8

REVIEW

Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis

Rose Hatala · David A. Cook · Benjamin Zendejas · Stanley J. Hamstra · Ryan Brydges

Received: 7 January 2013 / Accepted: 13 May 2013

Abstract Although feedback has been identified as a key instructional feature in simulation based medical education (SBME), we remain uncertain as to the magnitude of its effectiveness and the mechanisms by which it may be effective. We employed a meta-analysis and critical narrative synthesis to examine the effectiveness of feedback for SBME procedural skills training and to examine how it works in this context. Our results demonstrate that feedback is moderately effective during procedural skills training in SBME, with a pooled effect size favoring feedback for skill outcomes of 0.74 (95 % CI 0.38–1.09; $p < .001$). Terminal feedback appears more effective than concurrent feedback for novice learners' skill retention. Multiple sources of feedback, including instructor feedback, lead to short-term performance gains although data on long-term effects is lacking. The mechanism by which feedback may be operating is consistent with the guidance

R. Hatala
Department of Medicine, University of British Columbia, Vancouver, BC, Canada

R. Hatala (✉)
St. Paul's Hospital, Suite 5907, Burrard Bldg, 1081 Burrard St, Vancouver, BC V6Z 1Y6, Canada
e-mail: rhatala@mac.com

D. A. Cook
Office of Education Research, Mayo Medical School, Rochester, MN, USA

D. A. Cook
Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA

B. Zendejas
Department of Surgery, Mayo Clinic College of Medicine, Rochester, MN, USA

S. J. Hamstra
University of Ottawa Skills and Simulation Centre, Academy for Innovation in Medical Education, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

R. Brydges
Department of Medicine and The Wilson Centre, University of Toronto, Toronto, ON, Canada

hypothesis, with more research needed to examine other mechanisms such as cognitive load theory and social development theory.

Keywords Simulation-based medical education · Procedural skills training · Feedback · Motor learning · Technical skills

Introduction

Feedback to learners has long been recognized as a central component of effective clinical education (Ende 1983). Feedback provides learners with information about their performance with the goal of improving that performance. In a historical review of simulation-based medical education (SBME), feedback appeared as the most important and most frequently cited instructional design feature for effective SBME (McGaghie et al. 2010).

Procedural skills training is one area in which SBME plays a central role in the education of both novice and expert learners (Passiment et al. 2011). In many ways, the simulation setting is an optimal environment for providing feedback as learners can rehearse the key physical movements, patients are not at risk, training can be structured to optimize learning, and faculty are usually present to supervise and directly observe skill acquisition.

While previous meta-analyses have focused on the effectiveness of procedural skills training in SBME (Gurusamy et al. 2008; Ma et al. 2011; Sutherland et al. 2006), none have closely examined the specific role of feedback. We recently completed a large systematic review examining the general effectiveness of SBME. In two related reports, we used subgroup meta-analysis to explore the role of feedback and found inconsistent educational benefits across the various Kirkpatrick's outcomes (Kirkpatrick and Kirkpatrick 2009). Specifically, when SBME was compared to no-intervention, there was either no benefit to higher levels of feedback or the outcomes favoured lower amounts of feedback (Cook et al. 2011). When comparing SBME to other non-simulation interventions (Cook et al. 2013), higher levels of feedback were generally associated with larger effect sizes, although these were not always statistically significant. Such subgroup analyses evaluate between-study rather than within-study differences, which limits their usefulness. A third report in this series examined studies comparing simulation-based interventions and generally demonstrated larger effect sizes for higher levels of feedback although this association was only statistically significant for one of five outcomes (Cook et al. 2012). However, none of these studies focused specifically on procedural skills training. A systematic review of the SBME literature focused on procedural skills training, that concentrates on within-study variations in feedback by comparing two simulation-based interventions, could inform and deepen our current understanding of this instructional design feature.

The present review represents a planned sub-analysis of the original studies identified in the above-mentioned systematic review. Our aim is to extend our understanding of feedback *generally* by synthesizing the evidence related to feedback in the *focused field* of SBME procedural skills training. Purely quantitative synthesis of studies in complex educational systems would allow us to examine whether an intervention such as feedback works, but would not permit deeper understanding of the mechanisms that underlie this instructional design feature. Consequently, we also undertook a narrative synthesis of the primary studies, inspired by realist review methods (Pawson et al. 2005a), to examine the mechanisms by which feedback affects learning. By this combined approach, we aim to deepen our understanding not just of whether feedback works, but how, for whom, and in what circumstances.

Conceptual frameworks

Key to most realist reviews is consideration of relevant explanatory theories or frameworks that may be used to explain the mechanisms by which an intervention has its effects. After conducting a focused literature review, we selected the following frameworks given their potential application to understanding the mechanisms of feedback during procedural skills training.

One framework is described as the 'guidance hypothesis' in the motor learning literature. Evidence supporting this hypothesis suggests that constant feedback from an instructor during each practice attempt (concurrent feedback) may lead to an over-reliance on the feedback such that when feedback is withdrawn, the learner's performance declines (Schmidt and Lee 2011). The guidance hypothesis may depend on task difficulty and learner experience, as researchers have found that learners performing simple tasks benefit from feedback at the end of a practice attempt (terminal feedback) whereas learners performing more complex tasks are aided by concurrent feedback (Wulf and Shea 2004).

Cognitive load theory can also be used to understand feedback during procedural skills learning (Van Merriënboer and Sweller 2005). Cognitive load theorists assume that the human cognitive processing system has finite capacity and refer to any burden placed upon this system as 'cognitive load'. Feedback provided during a procedural skills session could influence cognitive load, either increasing it by providing 'information-overload', or decreasing it by structuring the task so that it is better understood.

Social development theory is also relevant and has been applied previously as a framework for surgical skills development (Dunphy and Dunphy 2003). Theorists describe learning as occurring in a sociocultural context in which everyone involved in the learning (i.e. the learner, their peers and instructors), their actions and the meaning assigned to those actions, play a role in the emerging cognitive development of the learner (Vygotsky 1978). Vygotsky postulated that between assisted and unassisted performance is the Zone of Proximal Development, within which the learner can make developmental progress in collaboration with others even though they would not have been able to do so independently. An instructor provides feedback and guidance to the learner while in this zone.

Historically, feedback in education has been found to both help and hinder learning. One-third of studies included in a meta-analysis of feedback in education, for example, demonstrated negative effects on learning (Kluger and DeNisi 1996). Regulatory focus theory is a framework that may permit study of the negative effects associated with feedback (Higgins 1997). According to this theory, two systems of self-regulation modulate an individual's motivation: self-regulation with a promotion focus (for tasks we want to do) versus self-regulation with a prevention focus (for tasks we have to do). A recent study based on this theory showed that clinical trainees' framework for self-regulation interacted with the feedback message to either help or hinder learning (Watling et al. 2012).

Methods

Questions

We sought to answer (1) how effective is feedback in facilitating SBME procedural skills learning? and (2) how does feedback work and which mechanisms have a positive effect on learning outcomes? We defined feedback as information on performance received from an instructor, a peer, or a computer, either during or after the simulation activity (high feedback = structured and substantive from more than one source, low = infrequent and unstructured) (Cook et al. 2011). We defined technology-enhanced simulation as an educational tool or device with which the learner physically interacts to mimic an aspect of clinical care. Studies that evaluated only computer-based virtual patients or standardized patients were excluded. Below, we outline the methods specific to the quantitative systematic review and the qualitative critical narrative synthesis.

Systematic review methods

The general methods for the systematic review have been previously reported (Cook et al. 2011) and are presented here in abbreviated format. This review was planned, conducted, and reported in adherence to PRISMA standards of quality for reporting meta-analyses (Moher et al. 2009). As no human subjects were involved, IRB approval was not required.

Study eligibility

Eligible studies included those published in any language that focused on feedback during technology-enhanced SBME procedural skills training of health professions learners, at any stage of their training or practice. We excluded single-group pretest–posttest studies. Further, we excluded studies that compared two different simulators or compared SBME to another educational intervention because the change in instructional modality constitutes a confounding variable that limits interpretation of the study results (Cook et al. 2008).

Study identification and selection

An exhaustive literature search was performed searching multiple databases using a strategy previously reported in detail (Cook et al. 2011). No beginning date was used, and the last date of search was May 11, 2011. We searched for additional studies in the reference lists of all included articles.

All titles, abstracts, and full text of potential articles were screened independently and in duplicate for inclusion.

Data extraction

Data were extracted independently and in duplicate for all variables. We extracted the training level of learners, clinical topic, training location, study design, method of group assignment, outcomes, and methodological quality of the studies [graded using the Medical Education Research Study Quality Instrument (MERSQI)] (Reed et al. 2007). We also identified studies that contrasted four pre-specified instructional conditions: (1) feedback versus no feedback, (2) instructor- versus simulator-provided feedback, (3) feedback from one versus multiple sources, and (4) the timing of feedback (concurrent with the task or at the end of the task).

Information was abstracted separately for all learning outcomes, categorized as knowledge, time, and skill measures (i.e., processes or outcomes in the simulation context). No studies assessed the impact of the intervention on behavior (i.e., in the clinical context) or patient outcomes. If multiple measures of a skill outcome were reported, we selected a single outcome using our previously established order of priority (Cook et al. 2011).

Quantitative data synthesis

For each of the four contrasts noted above, we planned to quantitatively synthesize results using meta-analysis whenever three or more studies met the inclusion criteria for the contrast. For each reported outcome we calculated the standardized mean difference (Hedges' g effect size) using accepted techniques (Borenstein 2009; Cook et al. 2011; Hunter and Schmidt 2004; Morris and DeShon 2002). We quantified the inconsistency (heterogeneity) across studies using the I^2 statistic (Higgins et al. 2003). Due to large inconsistency in our analyses, we used random effects models to pool weighted effect sizes. SAS 9.2 (SAS Institute, Cary, NC, USA) was used for all analyses. Statistical significance was defined by a two-sided alpha of 0.05. Educational significance was based on Cohen's effect size classifications (0.2–0.49 = small; 0.5–0.8 = moderate, and >0.8 = large) (Cohen 1988).

Narrative synthesis methods

The scope of our narrative synthesis was to examine the mechanisms that underlie effective feedback for procedural skills SBME, and to examine for whom feedback works and in what circumstances. We adapted realist review methods, as this approach provided a useful structure within which to conduct a critical narrative synthesis of the literature (Pawson et al. 2005a).

We examined the included articles to identify additional theories that relate to feedback and discovered attentional focus and motivational impact of feedback, which we added to the theories outlined in the introduction (Wulf and Shea 2004). We also identified augmented feedback (knowledge of performance versus knowledge of results) as another important motor learning theory (Wulf and Shea 2004).

We extracted additional information on the learners, objectives, and context that might explain convergent and divergent results between studies. Finally, we assessed how well the findings of each study aligned with the explanatory theories. Using this methodology, we selected and refined our list of theories iteratively to ensure they applied to the primary studies.

Results

General results

Trial flow

Using our search strategy we identified 10,297 articles with an additional 606 identified from our review of reference lists and journal indices. From these we identified 31 articles for inclusion in the present review (Fig. 1; Table 1).

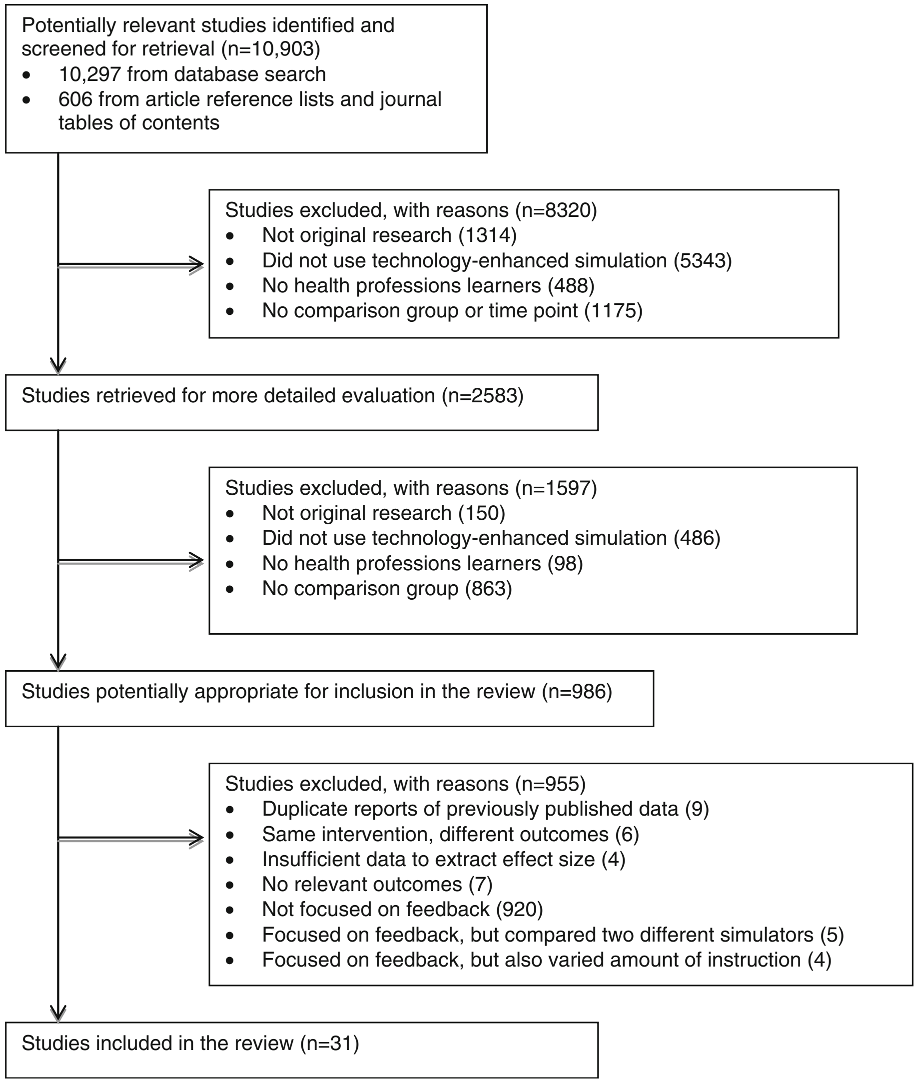

Fig. 1 Study flow

Study quality

Of the 31 included studies, 24 were randomized, controlled trials. Raters or outcomes were blinded in 22 studies. Ten studies lost more than 25 % of participants prior to outcome evaluation or failed to report follow-up. All outcomes were determined objectively. The mean study quality as assessed by MERSQI was 12.6 (SD 1.3) out of 18, which is higher than that in most medical education studies (Reed et al. 2007).

Table 1 Description of studies included in systematic review

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Kovacs et al. (2000)	Intubation	After instruction on advanced airway mannequin, provided with verbal performance feedback during two later evaluations	Same instruction, no feedback. Third arm with extra practice and feedback	Randomized, 3 arm	84 undergraduate MD students	Feedback improved performance	Feedback versus none
Pugh et al. (2001)	Pelvic examination	Part-task pelvic examination trainer with computer display of hand location and amount of pressure applied	Same part-task trainer without feedback component. Third arm without simulator training	Randomized, 3 arms	87 undergraduate MD students	Feedback improved performance	Feedback versus none
Byrne et al. (2002)	General anaesthetic with complications	Manikin for training. Watched video of performance after each scenario	Same trainer, without video review	Randomized	32 anaesthesia residents	Videotape did not improve performance	Feedback versus none
Wierinck et al. (2005)	Dental cavity preparation	Virtual reality (VR) cavity preparation, with visual display of VR metrics	Same trainer, without feedback. Third arm without training	Randomized, 3 arm	28 dental students	Feedback did not improve performance	Feedback versus none
Rosser et al. (2006)	Laparoscopic drills	Part-task laparoscopic drill trainer, with visual and auditory alarms	Same trainer, without alarms	Historical control	817 surgeons	Drills took longer with feedback	Feedback versus none (time outcome only)
Lazarski et al. (2007)	Mandibular repair	Part-task trainer for mandibular repair, with numerical score indicating force at which repair failed	Same trainer, without feedback. Third arm with instruction and feedback. Fourth arm of instruction without feedback	Cohort, 4 arms	61 undergraduate MD and dental students	Feedback improved performance	Feedback versus none

Table 1 continued

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Spooner et al. (2007)	Basic Life Support	Mannequin with continuous visual feedback on compression depth, hand position and ventilation volume	Same trainer, without feedback	Randomized	100 healthcare students	No difference between groups on global assessment	Feedback versus none
Rafiq et al. (2008)	Suturing+ knot tying	Part-task trainer, with visual feedback of forces generated during tasks	Same trainer, without feedback	Cohort	12 undergraduate MD students	Feedback improved performance	Feedback versus none
Chang et al. (2008)	Laparoscopic drills	Part-task laparoscopic drill trainer with feedback from instructor	Same instructional session, no feedback	Randomized	24 pairs of undergraduate MD students, junior surgical residents and non-medical personnel	Feedback improved performance	Feedback versus none
Day et al. (2009)	Tracheal suctioning	Instructional session with instructor and mannequin. Delayed written feedback at 7–10 weeks	Same instructional session, no feedback	Randomized	95 nurses and physiotherapists	Feedback improved knowledge and performance	Feedback versus none
Domuracki et al. (2009)	Cricoid pressure	Part-task trainer, with visual feedback of forces generated during task	Same trainer, without feedback	Randomized	101 undergraduate MD students and nurses	Feedback improved performance	Feedback versus none

Table 1 continued

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Moulton et al. (2009)	Communication skills during basic surgical procedures	Hybrid simulation with part-task trainer and standardized patient (SP). Verbal feedback from SP with video review	Same instructional session, no feedback	Randomized	32 undergraduate MD students and junior surgical residents	Feedback improved performance	Feedback versus none
Yasukawa (2009)	Dental cavity preparation	VR cavity preparation, with verbal feedback from instructor based on VR metrics	Same instructional session, no feedback	Randomized	39 dental students	Feedback improved time and performance	Feedback versus none
Gerling and Thomas (2003)	Clinical breast examination	Part-task trainer, with visual feedback of forces generated during task	Same instructional session, no feedback. Third arm, no training	Cohort, 3 arms	18 undergraduate MD students	Feedback improved performance	Feedback versus none
Chang et al. (2007)	Joint manipulation	Part-task trainer, with concurrent visual display of force feedback	Part-task trainer, with terminal visual display of force feedback. Third arm, same trainer without feedback	Randomized, 3 arms	36 undergraduate physiotherapy students	Feedback improved performance. Concurrent superior to terminal feedback	Feedback versus none; Concurrent versus Terminal
O'Connor et al. (2007)	Laparoscopic suturing	Part-task laparoscopic trainer with simulator-generated and expert performance feedback	Same trainer, simulator-generated feedback. Third arm, no feedback	Randomized, 3 arms	9 undergraduate MD students	Feedback improved performance. Reduced errors with multiple sources of feedback	Feedback versus none; Multiple versus single source

Table 1 continued

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Van Sickle et al. (2007)	Laparoscopic drill	Part-task laparoscopic drill trainer with auditory alarm	Same trainer, instructor served as alarm. Third arm with both auditory and instructor alarm. Fourth arm without feedback	Randomized, 4 arms	32 undergraduate MD and university students	Feedback improved performance. Highest performance with multiple sources of feedback	Feedback versus none; Instructor versus Simulator-generated; Multiple versus single source
Xeroulis et al. (2007)	Suturing and instrument knot tying	Part-task trainer, with concurrent feedback during each trial provided by expert	Part-task trainer, with terminal feedback after each trial provided by expert. Third arm, no feedback. Fourth arm, self-instruction	Randomized, 4 arm	60 undergraduate MD students	Feedback improved performance. Concurrent superior to terminal at immediate testing. Terminal superior to concurrent at 1-month retention test	Feedback versus none; Concurrent versus Terminal
Walsh et al. (2009)	Colonoscopy drills	Part-task trainer, with concurrent feedback during each trial provided by expert	Part-task trainer, with terminal feedback after each trial provided by expert	Randomized	30 undergraduate MD students	Groups similar at immediate and 1-week retention test. Terminal superior at transfer task	Concurrent versus Terminal
Quinn et al. (2003)	Cavity preparation	VR trainer with concurrent visual feedback	Same trainer, without VR feedback, but with instructor-generated feedback	Randomized	22 undergraduate dental students	Instructor-generated feedback group had higher performance	Instructor versus Simulator-generated
Enebo and Sherwood (2005)	Spine manipulation	Part-task trainer, with visual display of force feedback	Same trainer, without visual feedback, but with instructor-generated feedback	Randomized	33 chiropractic students	Improved performance with simulator-generated feedback at retention test	Instructor versus Simulator-generated

Table 1 continued

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Porte et al. (2007)	Knot tying and suturing	Part-task trainer with computer-generated feedback including proficiency criteria	Same trainer, expert feedback only. Third arm with computer-generated feedback, no proficiency criteria	Randomized, 3 arms	45 undergraduate MD students	Feedback improved performance. Instructor-generated feedback had higher performance at retention test	Instructor versus Simulator-generated
Wierinck et al. (2006a)	Cavity preparation	VR cavity preparation, with concurrent visual feedback	Same trainer, feedback during 2/3 trials. Third arm with no practice	Randomized, 3 arms	36 undergraduate dental students	No difference between groups	None
Stefanidis et al. (2007)	Laparoscopic suturing	Part-task laparoscopic trainer with concurrent expert feedback	Same trainer with concurrent expert feedback for 1/6 of session. Third arm with limited feedback but additional instructional video tutorial	Cohort	34 undergraduate MD students	Reduced frequency feedback group had shorter training to proficiency time	None
Rissanen et al. (2008)	Digital rectal examination	VR trainer with visual feedback display and expert feedback	Same trainer, with only visual feedback	Cohort	8 undergraduate MD students	Better performance with single source feedback	Multiple versus single source
Kruglikova et al. (2010)	Colonoscopy	VR colonoscopy, with VR generated metrics. Additional structured verbal feedback by instructor	Same instructional session, without instructor feedback	Randomized	22 residents	Instructor feedback group had improved time and performance	Multiple versus single source
Dine et al. (2008)	CPR	Mannequin with automated feedback messages from the defibrillator. Additional instructor-generated feedback	Same trainer, without automated feedback. Additional instructor-generated feedback	Randomized	65 nurses	Multiple sources of feedback improved performance	Multiple versus single source

Table 1 continued

Study, year	Procedure	Intervention	Comparator	Study design	Learner	Outcome	Quantitative analyses
Backstein et al. (2005)	Vascular anastomosis	Human, animate and inanimate models. Expert feedback during practice plus video review with expert	Same models and expert feedback, no video review	Randomized	26 surgical residents	Trend toward improved performance with multiple sources of feedback	Multiple versus single source
Scaringe (2002)	Spine manipulation	Part-task trainer, with concurrent visual display of force feedback. Instructor also commented on force	Same trainer, without visual feedback, but with instructor-generated feedback	Randomized	71 chiropractic students	No differences between groups	Multiple versus single source
Kahol et al. (2007)	Laparoscopic drills	VR trainer, with concurrent visual display of hand movements. Computer-generated feedback after practice trial	Same trainer, without concurrent feedback. Computer-generated feedback after practice trial	Cohort	8 residents	Both groups improved with feedback	None
Wierinck et al. (2006b)	Dental cavity preparation	VR cavity preparation, with visual display of VR metrics. 10 min tutorial with expert, prior to training	Same instructional session without tutorial. Third arm, without training	Randomized, 3 arms	36 dental students	Tutorial group had higher performance than feedback-only group	None

Study characteristics

Four of the studies had multiple arms, (Chang et al. 2007; O'Connor et al. 2007; Van Sickle et al. 2007; Xeroulis et al. 2007) with comparisons between different feedback interventions, such that the contrasts (i.e., discrete comparisons) described here total more than 31. There were 18 studies that compared a feedback intervention to a control intervention with no feedback (Byrne et al. 2002; Chang et al. 2007; Day et al. 2009; Domuracki et al. 2009; Gerling and Thomas 2003; Kovacs et al. 2000; Lazarski et al. 2007; Moulton et al. 2009; O'Connor et al. 2007; Pugh et al. 2001; Rafiq et al. 2008; Rosser et al. 2006; Chang et al. 2008; Spooner et al. 2007; Van Sickle et al. 2007; Wierinck et al. 2005; Xeroulis et al. 2007; Yasukawa 2009). Of the remaining included studies, three compared concurrent (feedback provided throughout each practice trial) to terminal feedback (feedback provided after each practice trial) (Chang et al. 2007; Walsh et al. 2009; Xeroulis et al. 2007) and two other studies examined reducing the frequency of concurrent feedback (Stefanidis et al. 2007; Wierinck et al. 2006a). Four studies compared instructor-generated feedback to feedback delivered by the simulator (Enebo and Sherwood 2005; Porte et al. 2007; Quinn et al. 2003; Van Sickle et al. 2007). Seven studies compared delivering feedback from one compared to multiple sources (O'Connor et al. 2007; Rissanen et al. 2008; Backstein et al. 2005; Dine et al. 2008; Kruglikova et al. 2010; Scaringe 2002; Van Sickle et al. 2007). Two studies did not fit into any of these classifications: one examined the addition of a tutorial prior to learning with feedback (Wierinck et al. 2006b) and the other compared combined concurrent and terminal feedback from the simulator to terminal feedback from the simulator (Kahol et al. 2007).

Medical students comprised the majority of learners though dental students, chiropractic students, physical therapy students, residents, surgeons, and nurses also participated (see Table 1). The majority of studies assessed skills as the outcome measure.

Systematic review

Quantitative synthesis: meta-analyses

Data were pooled for four separate meta-analyses (Figs. 2, 3, 4, 5). All meta-analyses demonstrated moderate to high inconsistency ($I^2 \geq 40$ %). For the meta-analysis of the 18 studies comparing feedback to no feedback, 5 studies assessed time outcomes in 936 learners. Seventeen studies assessed skill outcomes in 653 learners (Fig. 2). Meta-analysis demonstrated that the presence of feedback during SBME procedural skills training is associated with improved skill outcomes with a moderate and statistically significant effect size of 0.74 (95 % CI 0.38–1.09; $p < .001$; $I^2 = 87$ %) (Fig. 2). The association with time outcomes favored the use of feedback, but was small and statistically non-significant [pooled effect size of 0.32 (95 % CI −0.23 to 0.87; $p = .25$; $I^2 = 89$ %)].

The meta-analyses that focused on instructional design features of feedback are shown in Figs. 3, 4 and 5. There was no significant difference between concurrent and terminal feedback, for skill outcomes assessed immediately at the end of the intervention [effect size 0.36 favoring concurrent feedback (95 % CI −0.20 to 0.93; $p = .21$; $I^2 = 73$ %)] or assessed at a delayed retention test occurring at least 5 days post-training [effect size 0.08 favoring concurrent feedback (95 % CI −0.51 to 0.67; $p = .78$; $I^2 = 75$ %)] (Fig. 3a, b, respectively). There was a moderate, but statistically non-significant, effect of instructor-generated feedback compared to that generated by the simulator [effect size 0.74 favoring

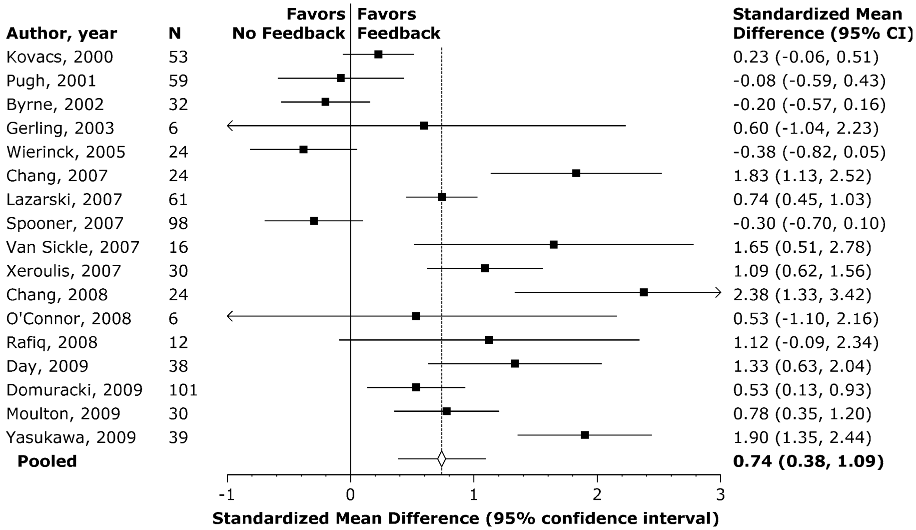

Fig. 2 Feedback versus no feedback, skill outcomes. Effect sizes for feedback compared with no feedback. Positive numbers favor feedback. For pooled effect size, $p < .001$; $I^2 = 87\ \%$

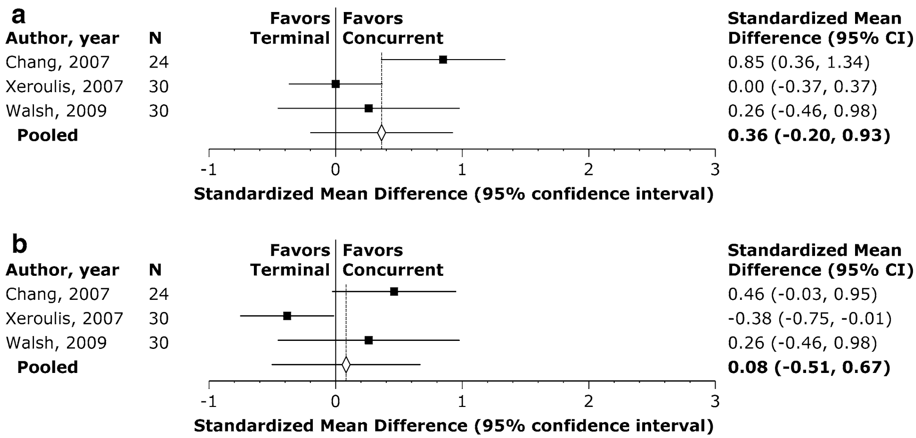

Fig. 3 **a** Concurrent versus terminal feedback, immediate post-test. Effect sizes for concurrent feedback compared to terminal feedback, for skills outcomes. Positive numbers favor concurrent feedback. For pooled effect size, $p = .21$; $I^2 = 73\ \%$. **b** Concurrent versus terminal feedback, delayed retention test. Effect sizes for concurrent feedback compared to terminal feedback, for skills outcomes tested at least 5 days after intervention. Positive numbers favor concurrent feedback. For pooled effect size, $p = .78$; $I^2 = 75\ \%$

instructor-generated feedback (95 % CI −0.23 to 1.71, $p = .13$, $I^2 = 88\ \%$)] (Fig. 4). Multiple sources of feedback compared to a single source of feedback led to enhanced learning outcomes with a small to moderate, statistically significant, effect size of 0.43 favoring multiple sources of feedback (95 % CI 0.07–0.79, $p = .02$, $I^2 = 42\ \%$) (Fig. 5).

Narrative synthesis

We have organized this synthesis according to the explanatory theories outlined previously.

Guidance hypothesis

Of the initial explanatory theories identified, the guidance hypothesis addresses the studies examining concurrent and terminal feedback. The guidance hypothesis suggests that provision of concurrent feedback may lead to an over-reliance on feedback resulting in a decrement in performance over the long-term (Wulf and Shea 2004). In the three studies that directly compared concurrent to terminal feedback during procedural skill acquisition (see Table 1 and Fig. 3) (Chang et al. 2007; Walsh et al. 2009; Xeroulis et al. 2007), learners' skills demonstrated improvement from pretest to immediate post-test after practicing in either feedback condition. There were no significant differences between the concurrent and terminal feedback groups at either immediate or short-term post-test (i.e. within 1 week of intervention), with the quantitative results suggesting a slight benefit to the concurrent feedback group (Fig. 3a, b). However, when the retention test occurred 1 month later (Xeroulis et al. 2007), or involved a transfer task test during the delayed post-test (Walsh et al. 2009), the terminal feedback group demonstrated superior procedural skills.

All three studies used novice learners and focused on teaching surgical procedural skills (knot-tying (Xeroulis et al. 2007)), colonoscopy manipulation (Walsh et al. 2009), and joint manipulation (Chang et al. 2007). In the two surgical studies (Walsh et al. 2009; Xeroulis et al. 2007), the experts providing the feedback were in the room, whereas feedback during joint manipulation consisted of a computer display of the force applied to the joint compared to a criterion force (Chang et al. 2007). Outcome assessment occurred immediately in all studies, with the addition of a retention test within 1 week (Chang et al. 2007; Walsh et al. 2009) or 1 month (Xeroulis et al. 2007) or a transfer test (i.e., task requiring learners to complete a novel, but related, simulator task) within 1 week (Walsh et al. 2009).

The pattern of results is consistent with the guidance hypothesis, in that learners in the concurrent feedback groups may have formed a reliance on feedback that supported short-term, but not long-term performance (Wulf and Shea 2004). This is especially apparent in Walsh et al.'s study, where the concurrent group had a significant decrement in performance on the transfer test compared to the post-test and retention test. Conversely, the terminal feedback group demonstrated sustained retention and transfer performance (Walsh et al. 2009).

Further supporting the guidance hypothesis, two studies examined the effect of reducing the frequency of feedback during a practice session (Stefanidis et al. 2007; Wierinck et al. 2006a). Both studies used novice learners, practicing laparoscopic suturing (Stefanidis et al. 2007) and dental cavity preparation (Wierinck et al. 2006a). In both studies, concurrent and terminal feedback was provided to one group at 100 % frequency, and at a reduced frequency for the comparison group; either for 10 of 60 min for medical students (Stefanidis et al. 2007) or for 2 of 3 practice trials for dental students (Wierinck et al. 2006a). Supporting the assertion that more feedback is not necessarily better, and consistent with the outcome postulated by the guidance hypothesis, performance did not differ between groups at immediate post-test in either study, and dental cavity preparation skills did not differ on 4 month retention and transfer tests.

Since only novice learners participated in the studies of feedback timing, we were unable to examine potential interactions between learner level and timing of feedback or between task complexity and timing of feedback, both of which have been studied in the guidance hypothesis literature (Wulf and Shea 2004). Outside of motor learning, feedback timing has been examined in the cognitive psychology literature particularly as it applies to testing effects on learning. Although the testing effect studies are not directly comparable to the studies in our review, the similarities in terms of the benefits of delayed feedback on long-term retention are worthy of further exploration (Roediger and Butler 2011).

Cognitive load theory

Unfortunately, only one study directly examined the cognitive load associated with a feedback intervention by specifically measuring learners' workload perceptions (O'Connor et al. 2007). The learners who received the most feedback (from both the expert and the simulator) during a laparoscopic suturing task rated their workload the lowest and demonstrated the highest performance outcomes. Without retention testing, it remains uncertain whether this is a long-term benefit.

We found indirect evidence that instructors may help to reduce the cognitive load of the learner by making the task more understandable. Specifically, studies in which instructors (rather than the simulator) provided feedback generally demonstrated that learners performed better with instructor feedback (Fig. 4). The benefits associated with instructor-derived feedback were found for novices learning simple tasks and also for residents learning the more complex surgical skills of colonoscopy (Kruglikova et al. 2010) and vascular anastomosis (Backstein et al. 2005). In addition, when learners received multiple sources of feedback (which always included instructor feedback) compared to a single source of feedback, higher learning outcomes were associated with multiple feedback sources (Fig. 5).

Explanatory theories that were unsupported

Many of our explanatory theories were not supported by the primary literature. None of the feedback studies specifically commented on social development theory as the mechanism by which feedback may have been exerting its effect, although one aligning finding is that interventions with instructor-based feedback were more effective than those with simulator-generated feedback alone. None of the studies explored the negative effects of feedback as postulated by the regulatory focus theory.

Aside from the timing of feedback examined under the guidance hypothesis, the content of the feedback (referred to as augmented feedback) is also important in the motor learning literature (Wulf and Shea 2004). Although we expected augmented feedback to be helpful in explaining the studies included in this review, there was a notable absence of either supportive or contradictory evidence in the SBME literature. Among the studies we identified, authors rarely distinguished between knowledge of performance and knowledge of results nor did they describe the content of the feedback message delivered by the instructor. Without additional SBME feedback studies specifically providing more description of the amount and type of feedback provided, this highly promising feedback theory from motor learning cannot be examined in the simulation literature.

We were also unable to determine if either learner motivation or attentional focus play a key role in effective feedback in the SBME context (Wulf et al. 2010). None of the studies

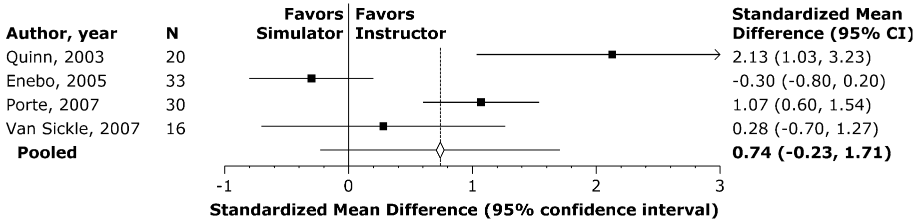

Fig. 4 Instructor-generated versus simulator-generated feedback. Effect sizes for instructor-generated versus simulator-generated feedback, for skills outcomes. Positive numbers favor instructor-generated feedback. For pooled effect size, $p = .13$, $I^2 = 88\ \%$

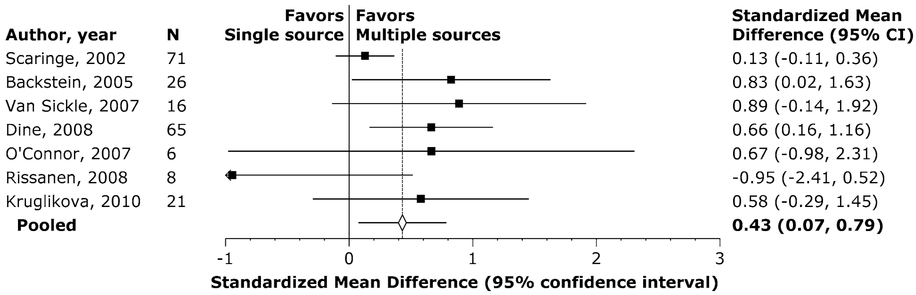

Fig. 5 Multiple sources of feedback versus a single source of feedback. Effect sizes for multiple sources of feedback versus a single source of feedback, for skills outcomes. Positive numbers favor multiple sources of feedback. For pooled effect size, $p = .02$, $I^2 = 42\ \%$

addressed the learners' motivation as impacted by feedback and nearly all failed to describe feedback in sufficient detail to determine the attentional focus of the learner.

Discussion

Our meta-analysis and narrative synthesis identifies several important messages regarding feedback for SBME procedural skills training. First, the meta-analytic results suggest a moderate benefit of feedback for SBME procedural-skills training. Regarding how to best implement feedback, terminal feedback may be better than concurrent feedback for long-term skill retention in novices learning simple tasks. Moreover, multiple sources of feedback (which typically include feedback from an instructor) appear superior to any single source of feedback when outcomes are assessed immediately after the intervention. Although this immediate benefit is notable, the studies focused on the source of feedback did not conduct delayed testing, which leaves us unable to determine if the apparent learning gains were sustained.

We also found evidence to support some potential mechanisms by which feedback may positively influence procedural skills training outcomes in SBME. The superiority of terminal feedback for novice long-term skill retention provides support for the guidance hypothesis (Wulf and Shea 2004). One study showed that learners had a lower perceived workload when presented with multiple sources of feedback, including instructor feedback, which may support cognitive load theory although additional focused research is necessary (van Merrienboer and Sweller 2010).

Strengths and limitations

The review includes a meta-analysis and a narrative synthesis in an attempt to capitalize on both quantitative and qualitative methodologies. The systematic review employed a comprehensive search strategy, rigorous data extraction and focused meta-analyses. The meta-analyses provide quantitative best estimates of effect for the design variations. The strengths of the narrative synthesis are the breadth of the explanatory theories and the iterative review process.

The limitations of this review parallel the limitations of the primary studies. Many studies do not describe in detail the amount and type of feedback provided, which constrains our exploration of how feedback may be effective. In addition, there is significant heterogeneity among the primary study results, limiting the interpretations of the pooled effect sizes. The narrative synthesis is also limited by the lack of theoretical grounding in many of the primary studies, leading us to rely on inferences with respect to explanatory theories. Finally, the paucity of delayed outcome assessments impairs our ability to evaluate the long-term effects of variations in feedback.

Integration with prior reviews

In considering how our findings intersect with other work, perhaps most relevant is a qualitative review from the general education literature, which focused on understanding how formative feedback works and developing guidelines relating to formative feedback (Shute 2008). Aligned with our findings, Shute describes formative feedback as decreasing the cognitive load for novice learners. She found conflicting literature regarding the complexity of the feedback, but found that the nature and quality of the feedback as it pertains to learning goals may be most important. As we have emphasized in this review, these general education findings support our assertion that in order for research results to be interpreted and meaningfully integrated, SBME investigators must present specific details regarding the nature, quality, complexity, and other features of the feedback provided to learners.

Within the medical education literature, a BEME review focused on assessment, feedback and clinical performance identified 13 previous feedback reviews as of 2003 (Veloski et al. 2006). None of these provided quantitative synthesis of the impact of feedback on performance, which we have been able to demonstrate within the SBME feedback for procedural skills literature.

Alternatively, several narrative reviews provide insights into feedback that are complementary to those in this review. In a 'state-of-the-science' article on effective feedback for health professional education, Archer provides a framework for feedback that investigators can use to describe key features of a feedback intervention, including focusing on the provision of feedback (type, structure and timing), the influence of the recipient (their receptivity to feedback and self-assessment capability and goal-setting) and the impacts of feedback on the learner and the organization (Archer 2010). While Archer is not prescriptive about how best to provide feedback, the emphasis on all of these elements echoes our own findings regarding some of the limitations of the primary studies reviewed here.

Focused more specifically on SBME, our results provide quantitative evidence in support of previous qualitative reviews that have identified feedback as an important instructional design feature (Issenberg et al. 2005; McGaghie et al. 2010). In addition to quantitative data, we add a more refined focus for researchers interested in feedback, particularly an emphasis on the timing of the message and its source.

Implications for practice and future research

The current review emphasizes what has often been stated without firm empiric support: feedback is effective for procedural skills training. Moreover, when teaching novice learners, the use of terminal feedback and a reduced frequency of feedback lead to effective learning and this aligns with much of the motor learning literature (Wulf and Shea 2004). Multiple sources of feedback, which typically include instructor feedback, also appear beneficial in the short-term but studies that confirm this by measuring long-term outcomes are necessary. Although our review focused exclusively on SBME procedural-skills training, it seems plausible that our results will generalize to patient-based procedural-skills training (Zendejas et al. 2013).

This area is ripe for well-designed qualitative and quantitative studies examining the optimal instructional design features for feedback in SBME procedural skills training. To date, many feedback studies lack a theoretical framework, making it difficult to ascertain how feedback may be operant. We need to draw on key theories from other disciplines, most notably motor learning, to see if and how they apply to SBME procedural skills. Among the theories with insufficient evidence to derive conclusions in this review, the role of augmented feedback and the influence of feedback on attention and motivation seem particularly promising as does the examination of cognitive load theory. Studies are needed with more advanced learners and more complex procedural skills, as some theories suggest that these factors may influence the mechanisms of feedback (Wulf and Shea 2004). Finally, it is important to examine learning effects using delayed retention tests, as what appear to be weaker immediate learning gains can translate into better long-term learning outcomes, as was demonstrated in the concurrent versus terminal feedback studies. Ultimately, new studies building upon the evidence summarized in this review will deepen our understanding of how best to use feedback in SBME procedural skills training and elucidate further how it works, for whom and in what circumstances.

Acknowledgments The authors thank Patricia Erwin, Jason Szostek, and Amy Wang for their assistance with literature searching, abstract reviewing and data extraction. This work was supported by intramural funds, including an award from the Division of General Internal Medicine, Mayo Clinic.

Conflicts of interest No relevant conflicts of interest for any of the authors. There was no industry relationship with this work.

References

Archer, J. C. (2010). State of the science in health professional education: Effective feedback. *Medical Education, 44*(1), 101–108.

Backstein, D., Agnidis, Z., Sadhu, R., & MacRae, H. (2005). Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery, 48*(3), 195–200.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY, US: Russell Sage Foundation.

Byrne, A. J., Sellen, A. J., Jones, J. G., Aitkenhead, A. R., Hussain, S., Gilder, F., et al. (2002). Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: A multicentre study. *Anaesthesia, 57*(2), 176–179.

Chang, J.-Y., Chang, G.-L., Chien, C.-J. C., Chung, K.-C., & Hsu, A.-T. (2007). Effectiveness of two forms of feedback on training of a joint mobilization skill by using a joint translation simulator. *Physical Therapy, 87*(4), 418–430.

Chang, S., Waid, E., Martinec, D. V., Zheng, B., & Swanstrom, L. L. (2008). Verbal communication improves laparoscopic team performance. *Surgical Innovation, 15*(2), 143–147.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Cook, D. A., Bordage, G., & Schmidt, H. G. (2008). Description, justification and clarification: A framework for classifying the purposes of research in medical education. *Medical Education, 42*(2), 128–133.
Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: A systematic review and meta-analysis. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare, 7*(5), 308–320.
Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher, 35*(1), e867-98.
Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *The Journal of the American Medical Association, 306*(9), 978–988.
Day, T., Iles, N., & Griffiths, P. (2009). Effect of performance feedback on tracheal suctioning knowledge and skills: Randomized controlled trial. *Journal of Advanced Nursing, 65*(7), 1423–1431.
Dine, C. J., Gersh, R. E., Leary, M., Riegel, B. J., Bellini, L. M., & Abella, B. S. (2008). Improving cardiopulmonary resuscitation quality and resuscitation training by combining audiovisual feedback and debriefing*. *Critical Care Medicine, 36*(10), 2817–2822.
Domuracki, K. J., Moule, C. J., Owen, H., Kostandoff, G., & Plummer, J. L. (2009). Learning on a simulator does transfer to clinical practice. *Resuscitation, 80*(3), 346–349.
Dunphy, B., & Dunphy, S. (2003). Assisted performance and the Zone of Proximal Development (ZPD); a potential framework for providing surgical education. *Australian Journal of Educational & Developmental Psychology, 3*, 48–58.
Ende, J. (1983). Feedback in clinical medical education. *The Journal of the American Medical Association, 250*, 777–781.
Enebo, B., & Sherwood, D. (2005). Experience and practice organization in learning a simulated high-velocity low-amplitude task. *Journal of Manipulative and Physiological Therapeutics, 28*(1), 33–43.
Gerling, G. J., & Thomas, G. W. (2003). Effect of augmented visual performance feedback on the effectiveness of clinical breast examination training with a dynamically configurable breast model. In *Systems, man and cybernetics, 2003. IEEE international conference on*, *3*, 2095–2100 vol. 3.
Gurusamy, K., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2008). Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *British Journal of Surgery, 95*(9), 1088–1097.
Higgins, E. T. (1997). Beyond pleasure and pain. *The American Psychologist, 52*(12), 1280–1300.
Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557–560. (Clinical research ed.).
Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27*(1), 10–28.
Kahol, K., French, J., McDaniel, T., Panchanathan, S., & Smith, M. (2007). Augmented virtual reality for laparoscopic surgical tool training. In *Proceedings of the 12th international conference on human–computer interaction: Applications and services*, pp. 459–467.
Kirkpatrick, D. L., & Kirkpatrick, J. D. (2009). *Evaluating training programs* (3rd ed.). San Francisco: Berrett-Koehler.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.
Kovacs, G., Bullock, G., Ackroyd-Stolarz, S., Cain, E., & Petrie, D. (2000). A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Annals of Emergency Medicine, 36*(4), 301–309.
Kruglikova, I., Grantcharov, T. P., Drewes, A. M., & Funch-Jensen, P. (2010). The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity virtual-reality simulation: A randomised controlled trial. *Gut, 59*(2), 181–185.

Lazarski, M. P., Susarla, S. M., Bennett, N. L., & Seldin, E. B. (2007). How do feedback and instructions affect the performance of a simulated surgical task? *Journal of Oral and Maxillofacial Surgery, 65*(6), 1155–1161.

Ma, I. W. Y., Brindle, M. E., Ronksley, P. E., Lorenzetti, D. L., Sauve, R. S., & Ghali, W. A. (2011). Use of simulation-based education to improve outcomes of central venous catheterization: A systematic review and meta-analysis. *Academic Medicine, 86*(9), 1137–1147.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education, 44*(1), 50–63.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., PRISMA Group. (2009, October). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*, 1006–1012.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105–125.

Moulton, C., Tabak, D., Kneebone, R., Nestel, D., MacRae, H., & LeBlanc, V. R. (2009). Teaching communication skills using the integrated procedural performance instrument (IPPI): A randomized controlled trial. *American Journal of Surgery, 197*(1), 113–118.

O'Connor, A., Schwaitzberg, S. D., & Cao, C. G. L. (2007). How much feedback is necessary for learning to suture? *Surgical Endoscopy, 22*(7), 1614–1619.

Passiment, M., Sacks, H., & Huang, G. (2011). *Medical simulation in medical education: Results of an AAMC Survey* (pp. 1–48). Washington, DC: Association of American Medical Colleges.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review—A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(Suppl 1), 21–34.

Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *American Journal of Surgery, 193*(1), 105–110.

Pugh, C. M., Srivastava, S., Shavelson, R., Walker, D., Cotner, T., Scarloss, B., et al. (2001). The effect of simulator use on learning and self-assessment: The case of Stanford University's E-Pelvis simulator. *Studies in Health Technology and Informatics, 81*, 396–400.

Quinn, F., Keogh, P., McDonald, A., & Hussey, D. (2003). A study comparing the effectiveness of conventional training and virtual reality simulation in the skills acquisition of junior dental students. *European Journal of Dental Education : Official Journal of the Association for Dental Education in Europe, 7*(4), 164–169.

Rafiq, A., Tamariz, F., Boanca, C., Lavrentyev, V., & Merrell, R. C. (2008). Objective assessment of training surgical skills using simulated tissue interface with real-time feedback. *Journal of Surgical Education, 65*(4), 270–274.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *The Journal of the American Medical Association, 298*(9), 1002–1009.

Rissanen, M. J., Kume, N., Kuroda, Y., Kuroda, T., Yoshimura, K., & Yoshihara, H. (2008). Asynchronous teaching of psychomotor skills through VR annotations: Evaluation in digital rectal examination. *Studies in Health Technology and Informatics, 132*, 411–416.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27.

Rosser, J. C, Jr, Colsant, B. J., Lynch, P. J., Herman, B., Klonsky, J., & Young, S. M. (2006). The use of a "hybrid" trainer in an established laparoscopic skills program. *Journal of the Society of Laparoendoscopic Surgeons, 10*(1), 4.

Scaringe, J. (2002). The effects of augmented sensory feedback precision on the acquisition and retention of a simulated chiropractic task. *Journal of Manipulative and Physiological Therapeutics, 25*(1), 34–41.

Schmidt, R., & Lee, T. (2011). *Motor control and learning: A behavioral emphasis* (5th ed.). Champaign: Human Kinetics.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Spooner, B. B., Fallaha, J. F., Kocierz, L., Smith, C. M., Smith, S. C. L., & Perkins, G. D. (2007). An evaluation of objective feedback in basic life support (BLS) training. *Resuscitation, 73*(3), 417–424.

Stefanidis, D., Korndorffer, J. R, Jr, Heniford, B. T., & Scott, D. J. (2007). Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. *Surgery, 142*(2), 202–206.

Sutherland, L. M., Middleton, P. F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., et al. (2006). Surgical simulation: A systematic review. *Annals of Surgery, 243*(3), 291–300.

van Merrienboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(1), 85–93.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177.

Van Sickle, K. R., Gallagher, A. G., & Smith, C. D. (2007). The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. *Surgical Endoscopy, 21*(2), 220–224.

Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A., & Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher, 28*(2), 117–128.

Vygotsky, L. (1978). *Mind in society: The development of higher mental processes*. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), (14th ed.). Harvard University Press.

Walsh, C. M., Ling, S. C., Wang, C. S., & Carnahan, H. (2009). Concurrent versus terminal feedback: It may be better to wait. *Academic Medicine : Journal of the Association of American Medical Colleges, 84*(10 Suppl), S54–S57.

Watling, C., Driessen, E., van der Vleuten, C. P. M., Vanstone, M., & Lingard, L. (2012). Understanding responses to feedback: The potential and limitations of regulatory focus theory. *Medical Education, 46*(6), 593–603.

Wierinck, E., Puttemans, V., Swinnen, S., & van Steenberghe, D. (2005). Effect of augmented visual feedback from a virtual reality simulation system on manual dexterity training. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe, 9*(1), 10–16.

Wierinck, E., Puttemans, V., & van Steenberghe, D. (2006a). Effect of reducing frequency of augmented feedback on manual dexterity training and its retention. *Journal of Dentistry, 34*(9), 641–647.

Wierinck, E., Puttemans, V., & van Steenberghe, D. (2006b). Effect of tutorial input in addition to augmented feedback on manual dexterity training and its retention. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe, 10*(1), 24–31.

Wulf, G., & Shea, C. (2004). Understanding the role of augmented feedback: The good, the bad, and the ugly. In A. Williams & N. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (1st ed., pp. 121–144). London: Routledge.

Wulf, G., Shea, C., & Lewthwaite, R. (2010). Motor skill learning and performance: A review of influential factors. *Medical Education, 44*(1), 75–84.

Xeroulis, G. J., Park, J., Moulton, C. A., Reznick, R. K., LeBlanc, V., & Dubrowski, A. (2007). Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery, 141*(4), 442–449.

Yasukawa, Y. (2009). The effectiveness of cavity preparation training using a virtual reality simulation system with or without feedback. *Kōkūbyō Gakkai zasshi. The Journal of the Stomatological Society, Japan*, *76*(2), 73–80.

Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient outcomes in simulation-based medical education: A systematic review. *Journal of General Internal Medicine*. doi:10.1007/s11606-012-2264-5