

Evidence-Based Effective Teaching Behaviors for Complex Psychomotor Skills Training

Alexander P. B. Alken^{1*}, Jan-Maarten Luursema¹, Lucas W. Thornblade²,
Xiaodong (Phoenix) Chen³, Louise Hull⁴, Karen Horvath², Harry Van Goor¹, Cornelia Fluit⁵

¹Department of Surgery, Radboud University Medical Center, Nijmegen, The Netherlands

²Department of Surgery, University of Washington, Seattle, USA

³Department of Surgery, Brigham and Women's Hospital, Boston, USA

⁴Centre for Implementation Science, King's College, London, UK

⁵Health Academy, Radboud University Medical Center, Nijmegen, The Netherlands

Email: *Sander.Alken@Radboudumc.nl

How to cite this paper: Alken, A. P. B., Luursema, J.-M., Thornblade, L. W., Chen, X. P., Hull, L., Horvath, K., Van Goor, H., & Fluit, C. (2019). Evidence-Based Effective Teaching Behaviors for Complex Psychomotor Skills Training. *Creative Education*, 10, 1285-1304.

<https://doi.org/10.4236/ce.2019.106097>

Received: May 5, 2019

Accepted: June 27, 2019

Published: June 30, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Introduction: Although the research to operating room teaching is extensive, evidence relating surgical teachers' behaviors to trainees' objective complex psychomotor skills acquisition is limited. We aimed to identify objectively evidence-based teaching behaviors in hands-on training associated with increased complex psychomotor skills in surgical and non-surgical trainees. **Methods:** The MEDLINE, PsycINFO and ERIC databases were searched for relevant papers. Due to comparable training characteristics to complex surgical skill acquisition, papers on sports and music training were also included. Paper screening took place after training sessions with the inclusion and exclusion criteria. Inter-rater reliability was determined. Data were extracted and the quality of studies was assessed with the Medical Education Research Study Quality Instrument (MERSQI) and the Newcastle-Ottawa Scale-Education (NOS-E). **Results:** 18,337 references were identified. Seven studies were included. Teaching behaviors shown to improve trainees' objective skills acquisition included feedback, instruction, active trainee involvement and demonstrations. Feedback and instruction with an external focus on the task and effect were supported by the strongest evidence. There was significant evidence regarding negative effects of harshly criticizing and belittling teaching behaviors. The data quality and evidence for most teaching behaviors were weak with low impact levels. **Discussion:** Feedback, instruction, active trainee involvement and demonstrations are important for the hands-on teaching of complex psychomotor skills. However, strong evidence supporting their effectiveness is lacking. Future research should be directed on investigating the relationship between clearly defined teaching behaviors and the objective acquisition of complex skills in surgical trainees.

Keywords

Surgical Education, Effective Intra-Operative Teaching, Surgical Skills Training

1. Introduction

Skills training is inseparably linked to surgical practice (Kneebone, 2003). Although simulation offers a safe training environment (Kneebone, 2003; Sutherland et al., 2006) teaching in the operating room (OR) remains central to surgical education (Agha & Fowler, 2015; De Win et al., 2016; Kneebone, 2003; Konge & Lonn, 2016; Li & George, 2017; Sutherland et al., 2006). While being taught by experienced surgeons, it is in the OR that trainees develop a set of skills, tacit knowledge, and abilities to cope with complex, stressful and variable situations (Agha & Fowler, 2015). Training trainees in the OR is a challenging task as teaching and safe surgical care have to be combined. To offer the best possible training it is important to understand how to successfully support learning in the OR. This may be even more important since the time and opportunities available for teaching in the OR are decreasing (Anderson et al., 2013; McCaskie, Kenny, & Deshmukh, 2011; Reznick & MacRae, 2006; Snyder et al., 2012).

The literature on teaching in the OR is extensive. Behaviors like providing feedback, instructions, explanations, demonstrations, asking questions, and using humor have been identified (Alken, Tan, Luursema, Fluit, & van Goor, 2013, 2015; Blom et al., 2007; Chen, Williams, Sanfey, & Smink, 2015; Hauge, Wanzek, & Godellas, 2001; Roberts, Brenner, Williams, Kim, & Dunnington, 2012; Sutkin, Littleton, & Kanter, 2015) and evaluated (Claridge et al., 2003; Jensen, Wright, Kim, Horvath, & Calhoun, 2012; Levinson, Barlin, Altman, & Satin, 2010; Rose, Waibel, & Schenarts, 2011). Novice trainees need these behaviors to be applied intensively, but as trainees progress, the intensity should be reduced and increasingly substituted by reflective feedback (White, Rodger, & Tang, 2016). Although many conceptual frameworks for teaching in the OR are available, most research relies on trainees' perceptions of satisfaction or expert opinions (McKendy et al., 2017). Research into what specific teaching behaviors contribute to the actual, objective acquisition of skills in trainees seems scarce. Studies on simulation training are also extensive (Dawe et al., 2014; Gurusamy, Aggarwal, Palanivelu, & Davidson, 2009; Sturm et al., 2008; Sutherland et al., 2006); however, it is again hardly addressed which specific hands-on teaching behaviors are used. This makes it difficult to draw firm conclusions and make evidence-based recommendations on how to teach in the OR from a behavior standpoint.

In order to address this evidence gap, we conducted a literature review study according to systematic review principles. Our aim was to identify hands-on, evidence-based teaching behaviors which have been shown to be effective for

complex psychomotor skills acquisition in adult trainees. We looked for teaching behaviors that were associated with an objectively measured improvement of psychomotor skills, not a perceived improvement. By gaining more knowledge on this topic and discussing our findings, we expect to be able to express recommendations for surgical teachers to help them teach complex surgical psychomotor skills.

2. Methods

2.1. Scoping Search

A scoping search showed that research on evidence-based teaching behaviors for hands-on surgical complex psychomotor skills was scarce. We considered a skill complex if it was executed with specialized equipment, involves multiple actions and requires conscious cognitive demands. Previous research on surgical skills teaching turned to the fields of sports and music education due to similarities in complexity and training intensity (McCaskie et al., 2011; White et al., 2016). Keeping with this approach, we searched for teaching behaviors in the fields of surgical, medical, sports, and music education.

2.2. Search Terms

Together with a librarian specialized in systematic reviews we defined our search strategy and searched studies published prior to February 14, 2018, in the MEDLINE, PsycINFO and ERIC databases (Table 1). All references were imported into Endnote X7 (Thompson Reuters, Philadelphia, PA, USA).

2.3. Screening

All references were screened on title and abstract according to a three-stage pre-screening and screening approach (Figure 1). For pre-screening, a medical student was trained using 3 sets of 100 references, independently screened by the student and one author (SA) achieving an inter-rater reliability of Cohen's Kappa 0.7. The student pre-screened all references using inclusion and exclusion criteria (Table 2). We took a conservative approach: in the case of any doubt, a reference was forwarded to the screening stage.

In the screening phase we applied stricter inclusion and exclusion criteria (Table 3). Two times 100 references were independently screened by two researchers (SA and JML) for training (Figure 1). An inter-rater reliability of Cohen's Kappa 0.7 was reached. All references were then independently screened by SA and JML, achieving a moderate inter-rater reliability of Cohen's Kappa 0.6. Disagreements were resolved by a third researcher (CF) who was informed about the disagreement, but who was blinded for the individual decisions made by the other researchers. Papers that were included by CF were accepted for full text screening.

For full text screening we applied the strict inclusion and exclusion criteria (Table 3). Due to the moderate inter-rater reliability obtained during screening,

two extra training rounds with a total of 14 papers selected by SA were organized to ensure agreement on the interpretation of the inclusion and exclusion criteria. All remaining papers were screened by SA on full text. CF and JML each screened half the amount. We only included papers in which the effectiveness of teaching behaviors was analyzed in relation to the objective acquisition of skills in trainees. Reference lists of included papers were screened for relevant references.

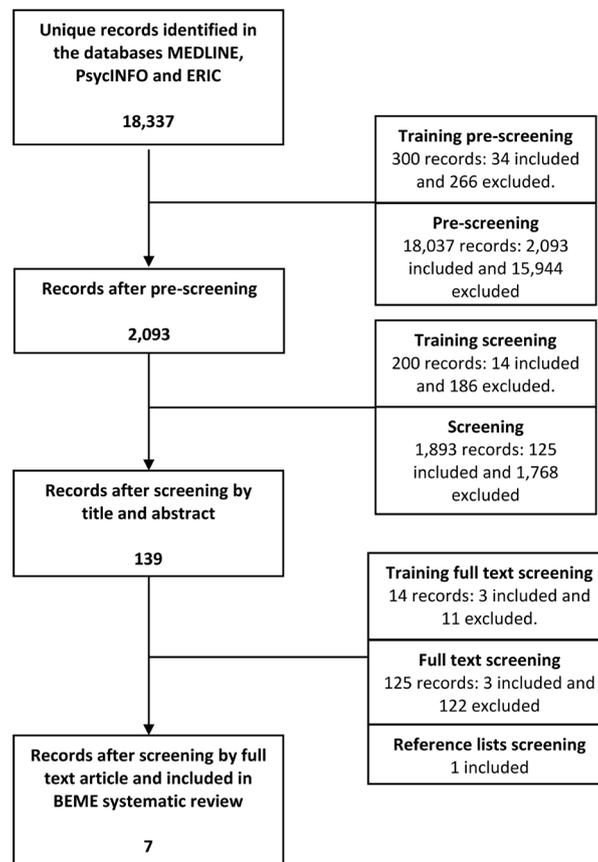


Figure 1. The selection of the reviewed articles.

Table 1. Search conducted in this systematic review study.

- 1) ((educat* or teach* or instruct* or tutor* or guid* or coach* or train* or feedback or mentor* or supervis* or pedagog* or faculty) adj3 (behavio* or activit* or practice* or interaction* or action* or characteristic* or strateg* or approach*)). mp. [mp = abstract, title, heading word, identifiers]
- 2) ((train* or student* or learn* or apprent* or intern or interns) adj3 (perform* or develop* or progress* or achiev* or competen* or advanc* or enhanc* or improv* or accomplish* or expert* or proficien* or skill* or gain* or grow* or outcome* or effect* or abilit*)). mp. [mp = abstract, title, heading word, identifiers]
- 3) (surg* or medic* or sport* or athlet* or music* or instrument* or hospital* or operating room* or operating theat* or intraoperative or intra operative). mp. [mp = abstract, title, heading word, identifiers]
- 4) 1 and 2 and 3
- 5) limit 4 to yr = "1995-Current"

Table 2. Inclusion and exclusion criteria used during the pre-screening stage.

<i>Inclusion</i>	1	Studies that investigate one or more teaching behaviors in hands-on training situations to assess the effects of these behaviors on the trainees' <i>complex psychomotor skills</i> * development.
<i>Exclusion</i>	1	Studies in which medical education, surgical education, music education or sports education is NOT the subject of research;
	2	Studies in which psychomotor motor skills are NOT the subject of research;
	3	Studies on the education, revalidation or training of clinical populations;
	4	Studies on non-adult populations;
	5	Studies which do not report original research (e.g. reviews, opinion pieces, letters to the editor, etc.);
	6	Studies published before 1995;
	7	Studies in other languages than English.

*We consider psychomotor skills to be complex if 1. specialized equipment is required for their execution, and or 2. dynamic decision making is needed to select the proper skill to execute, and or 3. execution of the skill requires conscious attention even after training (e.g., the skill is physically or cognitively challenging).

Table 3. Additional criteria for the screening stage on title and abstract and full text screening.

<i>Inclusion</i>	1	See Table 2 .
<i>Exclusion</i>	1	Studies that investigate the training of simple psychomotor skills;
	2	Studies which do NOT investigate the effects of teaching behaviors on the objective acquisition of skills in trainees;
	3	1-7 (Table 2).

2.4. Data Extraction, Quality Assessment and Level of Impact

For each experiment in each included study, we extracted general information (authors and field of research), study set-up (research aim and design), outcome measures (evidence-based outcome measures used in the study, data collection methods and bias risk assessment), and the teaching behaviors that were shown to be effective in the trainees' acquisition of skills.

To assess each experiment's quality and risk for bias we used the Medical Education Research Study Quality Instrument (MERSQI) and the Newcastle-Ottawa Scale-Education (NOS-E) for quantitative research (Cook & Reed, 2015). The MERSQI focuses on study design, number of institutions used for sampling, response rate, subjective or objective data collection, validity evidence of the applied instruments, appropriate data analyses, and impact of the outcome measures. The NOS-E focuses on representativeness of the trainees, selection and comparability of a comparison group, likeliness for study retention, and use of blinded assessors. Because of their different focus, the two scoring systems are considered complementary (Cook & Reed, 2015).

For each experiment, individual MERSQI and NOS-E items were scored and total scores were calculated (maximum MERSQI score: 18; maximum NOS-E score: 6), and compared to the normative score of 12.3 (MERSQI) and 3.58 (NOS-E) (Cook & Reed, 2015). If an experiment used two methods to analyze

data (two outcome measures), total MERSQI and NOS-E scores were given for each method and mean scores were calculated. All scores were assigned by one researcher (SA) and discussed with a second researcher (CF). A consensus process was applied in case of disagreement.

MERSQI and NOS-E total scores give an indication of the overall quality. However, quality assessment should also take into account the individual MERSQI and NOS-E items (Cook & Reed, 2015). Applying the MERSQI and NOS-E enabled us to compare and interpret the quality of studies in relation to the normative scores.

We independently determined the impact levels for each experiment based on recommendations for evidence in medical education from Belfield et al. (2001) who based their work on the research of Kirkpatrick (Belfield, Thomas, Bullock, Eynon, & Wall, 2001). We identified four main impact levels (Table 4).

3. Results

Of 18,337 references, seven studies describing eight experiments met the inclusion criteria (Table 3; Figure 1). Four were conducted in sports education (Harrison, Fellingham, Buck, & Pellett, 1995; Wulf, Lauterbach, & Toole, 1999; Wulf, McConnel, Gartner, & Schwarz, 2002), two in music education (Duke & Henninger, 1998; Henninger, Flowers, & Councill, 2006); one in critical care education (McSparron et al., 2015) and one in surgical education (Flinn et al., 2016). Data from all experiments were derived from quantitative research methods. Six experiments were entirely focused on adult teachers training adult trainees (Flinn et al., 2016; Harrison et al., 1995; Henninger et al., 2006; McSparron et al., 2015; Wulf et al., 1999; Wulf et al., 2002), while two included both adult and non-adult trainees (5th and 6th graders (Duke & Henninger, 1998); and 15 to 17 years old (Wulf et al., 2002)).

Table 4. Level of impact distinguished in our systematic review study.

1b. Reaction teachers/coaches—Opinions/satisfaction regarding the quality and effectiveness of the coaching behavior (e.g. teachers' responses on questionnaires after training)

1a. Reaction trainees—Opinions/satisfaction regarding the quality and effectiveness of the coaching behavior (e.g. trainees' responses on questionnaires after training)

2. Learner outcomes—Development of skills in learners (e.g. observation and assessment of newly trained skills acquired by trainees in an experimental setting)

3. Transfer to real practice—Development of skills in learners as they are applied in real practice (e.g. observations and assessment of newly trained skills in the real-life operating room)

4. Healthcare outcomes—Changes in health care outcomes that can be attributed to the development of skills in trainees (e.g. the complication rate in operating rooms where trainees with newly acquired skills are involved)

Impact levels are ranked from lowest (1) to highest (4). Studies with an impact level of 1a and 1b were excluded from our review study (see inclusion and exclusion criteria; Table 3).

3.1. Study Quality

Using the MERSQI and NOS-E checklists, we identified sources of bias per experiment (**Table 5**). Total MERSQI and mean total MERSQI scores (ranging from 11.5 - 15) were either near or above the normative score of 12.3 (Cook & Reed, 2015). Total NOS-E and mean total NOS-E scores for all experiments were below the normative score of 3.58 (ranging from 1 - 3.5). Only one experiment used a method which achieved a NOS-E score above the normative score (4 (Wulf et al., 2002)). The level of impact for all experiments was focused on learner outcomes (**Table 4** and **Table 5**). This enabled us to assess each experiment's quality and interpret the strength of the results.

3.2. Research in Surgical and Medical Skills Training

Flinn et al. (2016) compared the effects of four feedback behaviors on the objective acquisition of complex laparoscopic skills (**Table 5**). Trainees who received harshly criticizing feedback performed worse than trainees who received encouraging positive feedback, minimal and neutral feedback, and no feedback. Trainees who received encouraging positive feedback did not perform better than trainees who received minimal and neutral feedback, and no feedback. The researchers concluded that it was not positive feedback which improved learning, but harshly negative and threatening feedback which impaired learning.

McSparron et al. (2015) analyzed feedback and instruction behaviors of teachers while they were teaching subclavian central venous catheter (S-CVC) insertion to a trainee who was instructed to show challenging learning behavior. Subsequently, the teachers' feedback and instruction behaviors were related to the performance of real novice trainees in a consequent training session (**Table 5**). Positive feedback (interpreted by the researchers as constructive feedback), suggestions as to how to improve, and step-by-step demonstrations were positively correlated. The regular repetition of learning goals was negatively correlated. The researchers concluded that this may be less effective for technical skills teaching.

3.3. Research in Sports and Musical Skills Training

Duke and Henninger (1998) compared two types of feedback in teaching a musical performance skill: corrections focusing on what trainees had to improve versus corrections focusing on what trainees had done wrong (**Table 5**). No significant differences were found in trainees' acquisition of skills. The researchers concluded that neutrally provided improvement feedback and fault-focused feedback were equally effective.

Wulf et al. (1999, 2002) compared the effectiveness of two types of feedback and instructions in teaching sports skills, which either externally focused on the task and effect, or internally focused on how to move (**Table 5**). Regarding accuracy of the trained sports skills, trainees whose teacher provided externally focused feedback and instructions performed better during training and reten-

tion, regardless of trainee experience, and especially if teachers had provided feedback after every trial. Trainees could also see themselves how accurate they had performed (trainees could derive perceptual feedback). For movement quality, internally focused feedback and instructions were more effective, but only in novice trainees.

Henninger et al. (2006) compared the teaching behaviors of music teachers in teaching a wind instrument (Table 5). Active trainee involvement, by stimulating trainees to ask questions, think aloud, etc., was shown to be important to teach effectively.

Harrison et al. (1995) compared the effectiveness of volleyball training in which the trainees determined the sequence and speed of progress versus volleyball training in which the teacher made all decisions (Table 5). No clear conclusions on what was more effective could be drawn.

3.4. Strength of the Evidence

The evidence for the described teaching behaviors is mostly weak and limited to improved learner outcomes in a training setting (second lowest level of impact; Table 4). We scored the strength of evidence weak for the results and conclusions drawn by Harrison et al (1995), Duke and Henninger (1998), Henninger et al. (2006), McSparron et al. (2015) and Flinn et al. (2016). Table 5 shows detailed information regarding the MERSQI and NOS-E bias risk assessment.

We scored the strength of the studies conducted by Wulf et al. (1999, 2002) moderate, providing the strongest evidence of all studies included in this review. The mean total MERSQI scores are all above the normative score, and the highest of all experiments included in this review. The mean total NOS-E scores are all around, but mostly below the normative score (Table 5). Still, important risks for bias remain.

4. Discussion

Our goal was to identify evidence-based teaching behaviors which improved complex psychomotor skills acquisition in a hands-on training setting, applicable to the OR. Although research was abundant, only very few studies analyzed the effectiveness of teaching behaviors in relation to objective improvement of skills in trainees. Of 18,337 studies only seven studies met the inclusion criteria. Only one study focused on surgical skills teaching (in a simulation setting). Most evidence was derived from studies on teaching sports which, however, contains comparable complex psychomotor skills. We identified evidence-based effective teaching behaviors related to feedback, instruction and active trainee involvement.

Feedback and instructions were effective if provided non-threatening (Flinn et al., 2016) and positive (McSparron et al., 2015), if they contained suggestions as to how to improve and step-by-step demonstrations (McSparron et al., 2015), and if they made trainees externally focus on the task and effect (Wulf et al.,

1999; Wulf et al., 2002).

Threatening feedback was found to be harmful to trainees' skills acquisition (Flinn et al., 2016). The importance of non-threatening feedback is supported by surgical review studies (McKendy et al., 2017; Timberlake, Mayo, Scott, Weis, & Gardner, 2017). Threatening feedback causes stress in trainees (Flinn et al., 2016), which is considered harmful since it is induced by a factor outside the learning task itself (Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012; Vogel & Schwabe, 2016), namely the teacher. Interestingly, Flinn et al. (2016) did not find any effects of positive feedback, while McSparron et al. (2015) did find an increase in trainees' skills. Comparing both studies is difficult since exact definitions of positive feedback remained unclear. Duke and Henninger (1998) reported that negatively formulated feedback (focused on faults) and positively formulated feedback (focused on how to improve) were equally effective if provided with a neutral tone of voice. Suggestions as to how to improve were also shown to be effective (McSparron et al., 2015). This is supported by review studies on intra-operative teaching (McKendy et al., 2017; Timberlake et al., 2017). Whether feedback is effective depends on how it is provided, but also how it is received and interpreted (Ridder, McGaghie, Stokking, & Cate, 2015). This is a complex process influenced by variables like tone of voice, intention, non-verbal behavior and prior relationship attributes established. It would be interesting to investigate how these variables exactly relate to the content of feedback messages and the effects on skills acquisition.

Externally focused feedback and instructions on task and effect were superior to improve accuracy and movement quality in sport skills (Wulf et al., 1999; Wulf et al., 2002), especially if intensely provided in trainees with experience, and in a setting of perceptual feedback (Wulf et al., 2002). Only novice trainees improved movement quality more with internally focused feedback and instructions on how to move. In surgical teaching, intensely provided feedback and instructions are considered only effective in novice trainees (McKendy et al., 2017; Timberlake et al., 2017) and if trainees cannot derive sufficient perceptual feedback (Nicholls, Sweet, Muller, & Hyett, 2016; White et al., 2016). The finding that intense externally focused feedback and instructions are effective in trainees with experience and possibilities for perceptual feedback, may shed new light on the role of the focus, content and intensity of instructions and feedback, and the effectiveness for teaching different surgical (sub) skills in relation to trainee experience and perceptual feedback.

Step-by-step demonstrations have also been shown to improve skills (McSparron et al., 2015). The importance of deconstructing skills into a series of small steps is also stressed in surgical reviews (McKendy et al., 2017; Nicholls et al., 2016; Timberlake et al., 2017). It is considered important to only provide trainees with the key instructions as to what to do. Other information may cause cognitive overload and prevent learning through distraction (Nicholls et al.,

2016).

Henninger et al. (2006) found that actively involve trainees was effective in skills training. Although active involvement was not clearly defined, they considered it important for teachers to know when to direct trainees, when to allow them to talk, ask questions and verbalize actions and thoughts. Surgical review studies support this finding, although their evidence is primarily based on perceptions and not objective measurements (McKendy et al., 2017; Timberlake et al., 2017). Verbalization by trainees is also considered a key step in surgical skills training because gaining insight in trainees' reasoning processes helps to teach effectively (Nicholls et al., 2016).

Interestingly, the Peyton approach (Nicholls et al., 2016), which is often used in surgical skills training, seems to be compatible with the identified evidence-based teaching behaviors: non-threatening, externally focused feedback on task and outcome, instructions, suggestions how to improve, step-by-step demonstrations and active trainee involvement. The effectiveness of this Peyton approach may be improved by the integration of these behaviors. However, the Peyton approach requires one skill to be taught and performed at least four times in a row (Nicholls et al., 2016), which, in our view, questions its compatibility to teaching in the OR.

4.1. Strengths and Weaknesses of Our Study

The strength of our review is the focus on the objective measurement of effects of teaching behaviors on the acquisition of complex psychomotor skills. Our review addresses the growing need for optimal OR teaching behaviors as well as objective assessment of training quality and trainees' skill level to assure safe and effective surgery performed by surgical residents. Our research underlines that much can be gained in the field of surgical educational research. Since only one surgical paper was included one might question the applied inclusion and exclusion criteria we applied. However, the small amount of papers in comparable professions and sports may imply that attention for objective outcomes of different teaching behaviors is generally limited.

The teaching behaviors with the strongest evidence originated from sports research, which questions the translatability of our findings to surgery. Differences exist regarding very fine motor skills training (fingers). However, the intensity and extensity of training necessary to reach proficiency is comparable. Also, proven effective methods in sports, like mental skills training in athletes, have been shown applicable to surgery (Anton & Stefanidis, 2016; Rogers, 2006; White et al., 2016). A limitation of our review study is the use of the MERSQI and NOS-E instruments which are specifically designed for medical educational research. The items on these instruments, however, were formulated in a way that made them easily applicable to other research fields. This review only selected papers written in English. To our best estimation we excluded only seven out of 18,337 papers because they were written in non-English languages. This

makes us believe all relevant research is represented.

4.2. Implications for Future Research and Surgical Practice

We were surprised to find a lack of research investigating the effects of teaching behaviors on the objective acquisition of surgical skills. Future research should focus on defining teaching behaviors that are effective for the objective development of complex surgical skills in hands-on training settings (measuring impact beyond perceptions of effectiveness). We advise to study the possibilities for and frequency of externally focused feedback and instructions in surgical skills training and the effectiveness of clearly defined feedback messages in respect of differences regarding formulation, tone of voice, intention and non-verbal behavior. Controlled longitudinal studies are required, both conducted in simulated settings with possibilities of standardizing tasks and teaching behaviors, but also in the clinical real-life OR setting. We advocate to start the incorporation of externally focused teaching behaviors into surgical teaching practice, first in simulation training of simple skills, gradually followed by more complex skills in the real-life OR environment.

Table 5. Summary and bias risk assessment of the experiments which were included in our review study.

Authors	Field	Aim	Design	Evidence based outcome of learning	Bias risk assessment based on individual MERSQI and NOS-E items (Granted score per item/Maximum potential score per item)	MERSQI total score	NOS-E total score	Impact level
Duke & Henninger 1998	Music education <i>Soprano recorder</i>	To compare the effects of improvement focused versus fault focused corrective feedback on learning a music performance skill on a soprano recorder in elementary school and college students.	Non-randomized 2 group study <u>Setting:</u> Half the trainees received improvement and half received fault focused corrective feedback. One of the researchers provided all feedback and intended to keep a neutral tone. <u>Data collection:</u> The researchers assessed the skills acquisition of trainees based on video-recordings through the amount of trials and time a trainee needed to independently perform in a duet. They also observed and noted the amount of corrections on what to improve,	Trainees in the improvement focused corrective feedback condition did not differ from trainees in the fault focused corrective feedback condition regarding number of trials and time in minutes { A } needed to perform independently. Observations of the applied teaching behavior (classification of behaviors in 3 predetermined categories by the researchers { B }) showed that improvement focused corrective feedback was higher in the improvement focused condition, fault focused corrective feedback was higher in the fault focused	A: <u>MERSQI:</u> Non-randomized 2 group (2/3); Two institutions (1/1.5); Sampling response rate $\geq 75\%$ (1.5/1.5); Objective data assessment (3/3); Validity evidence N/A; Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). NOS-E: No representative group of participants (not randomly selected; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).	A: 14.4* B: 12 M: 13.2	A: 3 B: 3 M: 3	2

			corrections on what went wrong and the amount of positive feedback that was provided.	condition, but that the amount of positive feedback was equal in both conditions.	B: MERSQI: Non-randomized 2 group (2/3); Two institutions (1/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); No validity evidence established (0/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). NOS-E: No representative group of participants (not randomly selected; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).	
Flinn et al. 2016	Surgical education <i>Laparoscopic cutting task</i>	To analyze whether an encouraging feedback style results in less subjective stress and better laparoscopic skills in inexperienced medical students than a harshly criticizing feedback style, a quiet observational feedback style or no feedback.	Four group pre and post test (for the results relevant to our study) Setting: Trainees were equally divided into four groups during simulation training, based on order of sign up: a control group (with no feedback); an observation only group (teacher only observing the trainee using minimal and neutral feedback); an encouraging feedback style group (teacher providing positive feedback and acting encouraging), and a harshly criticizing feedback style (teacher providing harsh criticism, being sarcastic and condescending). Trainees were encouraged to perform as quickly and accurate as possible. Trainees	Trainees in all groups improved significantly during training. Trainees who were provided with harshly criticizing feedback behaviors scored lower than the trainees in the other groups on overall performance scores (time and accuracy scores transformed into a performance score based on the SAGES FLS scoring system; no further description { A }). Trainees provided with positive and encouraging feedback scored similar to trainees who received minimal and neutral feedback, and trainees who received no feedback.	A: MERSQI: Single group pre test post test (1.5/3); One institution (0.5/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); Validity evidence N/A; Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). NOS-E: No representative group of participants (sampling not described; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).	A: 13.2* A: 3 2

			<p>were naive to the study purpose.</p> <p><u>Data collection:</u> The time to task completion was recorded and accuracy scores were collected and transformed into performance scores (based on the SAGES FLS scoring system; no further description). Performance scores were compared for the trainees' first trial, last trial, and the mean performance score of all trials (ranging from 4 to 8) in between.</p>					
Harrison et al. 1995	Sport education <i>Volleyball</i>	To analyze whether a practice or command coaching style results in better volleyball sub skills performance in low, medium and highly skilled trainees.	<p>Non-randomized 3 group cohort study</p> <p><u>Setting:</u> Based on a skills pre-test (unclear who took this) trainees were divided in high, medium and low skilled. Training lasted for 19 days. Per skills group, 50% was provided with the command and 50% with the practice coaching style by a professor in Physical education (no further description).</p> <p><u>Data collection:</u> Data was collected on the pre-test, day 2 and 3 (midterm test), day 8 and 9 (midterm test) and day 17, 18 and 19 (post-test) according to the American Alliance for Health, Physical Education, Recreation and Dance test and the Stanley spike test (ratio correct to incorrect actions; no further description). Trainee performance was videotaped. Four observers (no further description) rated game play and practice trials.</p>	<p>Regarding practice trials, low skilled trainees in the command style condition performed better on volleyball set skills over the course of 19 days (ratio correct to incorrect performances {A}) and low skilled trainees in the practice style condition performed better on the volleyball spike skills (ratio correct to incorrect performances {B}).</p>	<p>A /B: MERSQI: Cohort study (2/3); One institution (0.5/1.5); Sampling response rate $\geq 75\%$ (1.5/1.5); Objective data assessment (3/3); Content validity and internal validity established, no relationships to other variables (2/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described; 0/1); Comparable control group (0/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).</p>	<p>A: 13.5 B: 13.5 M: 13.5</p>	<p>A: 2 B: 2 M: 2</p>	2

Henninger et al. 2006	Music education <i>Wind instrument</i>	To determine whether experienced teachers and inexperienced pre-service teachers have different effects on trainee progress, trainee performance and apply different teaching behaviors when teaching to play a wind instrument.	Single group post test (for the results relevant to our study) <u>Setting:</u> Trainees selected by the teachers were trained. Training was recorded on video. <u>Data collection:</u> Trainee performance was analyzed by independent experts on a 5-point Likert scale. Trainees were then ranked into the 33% best, intermediate and lowest performing trainees. Teacher and trainee behaviors were analyzed by the researchers with an observational instrument.	Trainees whose performance was classified higher (independent observer assessment: low, intermediate, high, calculated according to a 1 item 5-point rating scale score given by independent observers {A}) talked more during training (scoring on earlier developed observational instrument by researchers {B}). No significant pair-wise effects; no significant correlation.	<p>A: <u>MERSQI:</u> Single group post test (1/3); Three institutions or more (1.5/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); No validity evidence established (incorrect method for internal validity applied) (0/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (0/1); No comparable group (no separate comparison group; no intervention: 0/1); Non-comparable comparison group (no separate comparison group; no intervention: 0/1); Study retention unlikely source of bias (1/1); Blinded assessment (1/1).</p> <p>B: <u>MERSQI:</u> Single group post test (1/3); Three institutions or more (1.5/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); Content validity established, no internal validity (incorrect method applied) and relationships to other variables (1/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (0/1); No comparable group (no separate comparison group; no intervention: 0/1); Non-comparable comparison group (no separate comparison</p>	A: 11.5 B: 12.5 M: 12	A: 2 B: 1 M: 1.5	2
-----------------------	---	--	--	---	---	--	---	----------

					group; no intervention: 0/1); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).			
McSparron et al. 2015	Critical medical care education <i>Placing a subclavian central venous catheter (s-CVC)</i>	To develop a procedural objective structured teaching exercise (PrOSTE) and analyze procedural teaching behavior which improves trainee performance.	Single group post test <u>Setting:</u> A standardized trainee instructed to show challenging learning behavior was trained in s-CVC on a simulator. Consequently, in a following training session, groups of three novice trainees were trained in s-CVC on a simulator. <u>Data collection:</u> A standardized trainee rated the teaching skills of each teacher on a questionnaire. One of the researchers assessed the novice trainees' s-CVC skills on a checklist.	Higher ratings of a standardized trainee on 3 items (26-item trainee PrOSTE questionnaire; 1-5 { A }) related to higher trainee performance in a consequent training session of novice trainees (assessment by trained observer (researcher); S-CVC 34 item questionnaire; 1-5 { B }). The opposite was found for 1 item. No overall correlation between A and B was found.	A/B: <u>MERSQI:</u> Single group pre test post test (1/3); One institution (0.5/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); Content validity established, no internal validity and relationships to other variables (1/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described: 0/1); No comparable group (no separate comparison group; no intervention: 0/1); Non-comparable comparison group (no separate comparison group; no intervention: 0/1); Study retention unlikely source of bias (1/1); No blinded assessment (blinding not described; 0/1).	A: 11.5 B: 11.5 M: 11.5	A: 1 B: 1 M: 1	2
Wulf et al. 1999	Sports education <i>Golf putt</i>	To analyze and compare the effects of externally and internally focused instruction on learning the golf putt in trainees with no experience.	Randomized controlled trial <u>Setting:</u> Half the trainees received externally and half received internally focused feedback before each set of training trials. One of the researchers provided all instructions for each trainee during training. During retention no instructions were provided. <u>Data collection:</u> One	Trainees with externally focused instruction outperformed trainees with internally focused instruction on accuracy (hitting target areas with predetermined points, given by one of the researchers; 0-5 { A }) during training and retention.	A: <u>MERSQI:</u> Randomized controlled trial (3/3); One institution (0.5/1.5); Sampling response rate \geq 75% (1.5/1.5); Objective data assessment (3/3); Validity evidence N/A; Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described; 0/1); Comparable control group (1/1); Allocation of	A: 15*	A: 3	2

			of the researchers scored the trainees' accuracy scores based on hitting target areas.		trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).			
Wulf et al. 2002	Sports education Exp. 1 <i>Volleyball serve</i>	To analyze and compare the effects of externally and internally focused feedback on learning the volleyball serve in novice and advanced trainees.	4 group cohort study <u>Setting:</u> Half of the novice trainees received externally and half of the novice trainees received internally focused feedback. The same conditions were applied for the advanced trainees. All feedback was standardized, neutrally formulated and provided during training after every fifth trial (50 in total) by one of the researchers. During the retention test no feedback was provided. <u>Data collection:</u> Based on video recordings two independent observers gave movement quality scores according to guidelines.	Trainees with externally focused feedback after each set of trials during training outperformed trainees with internally focused feedback on accuracy (hitting target areas with predetermined points, given by one of the researchers; 0-4 { A }) during training and retention. Effects were independent of trainee experience. Trainees with externally focused feedback outperformed trainees with internally focused feedback on movement quality (independent observer scoring on observational instrument based on guidelines; 0-15 { B }) during training, regardless experience. During retention, the advanced externally feedback group outperformed the advanced internally feedback group. The novice internally feedback group outperformed the novice externally feedback group.	A: <u>MERSQI:</u> Cohort study (2/3); Two institutions (1/1.5); Sampling response rate $\geq 75\%$ (1.5/1.5); Objective data assessment (3/3); Validity evidence N/A; Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1). B: <u>MERSQI:</u> Cohort study (2/3); Two institutions (1/1.5); Sampling response rate $\geq 75\%$ (1.5/1.5); Objective data assessment (3/3); Content and internal validity established, no relationships to other variables (2/3); Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); Blinded assessment (1/1).	A: 14.4* B: 14 M: 14.2	A: 3 B: 4 M: 3.5	2

Wulf et al. 2002	Sports education Exp. 2 Soccer pass	To analyze and compare the effects of externally and internally focused feedback on learning the soccer pass with intermittent or constant feedback.	Randomized controlled trial <u>Setting:</u> Trainees all had experience, received either internally or externally focused feedback, either constant (after every trial) or intermittent (after every third trial; 30 in total). All feedback was standardized, neutrally formulated and provided by one of the researchers. During retention no feedback was provided. <u>Data collection:</u> Accuracy scores were noted by one of the researchers based on target areas.	The externally focused feedback group outperformed the internally focused feedback group on accuracy (hitting target areas with predetermined points, given by one of the researchers; 0-3 { A }) during training and retention, regardless feedback intensity. The intermittent internally focused feedback group outperformed the constant internally focused feedback group on accuracy during training and retention. The constant externally focused feedback group outperformed the intermittent externally focused feedback group.	A: <u>MERSQI:</u> Randomized controlled trial (3/3); One institution (0.5/1.5); Sampling response rate $\geq 75\%$ (1.5/1.5); Objective data assessment (3/3); Validity evidence N/A; Data analysis beyond descriptive (2/2); Data analysis appropriate (1/1); Outcomes measuring skills acquisition (1.5/3). <u>NOS-E:</u> No representative group of participants (sampling not described; 0/1); Comparable control group (1/1); Allocation of trainees not concealed (1/2); Study retention unlikely source of bias (1/1); No blinded assessment (0/1).	A: 15* A: 3 2
------------------	-------------------------------------	--	--	--	--	------------------------------------

M = The mean MERSQI and NOS-E total scores: $(A + B) / 2$.

* = In case instrument validity was not applicable and the maximum potential MERSQI score was 15, the MERSQI score was corrected and multiplied with a factor 1.2 in order to obtain a potential MERSQI total score of 18.

Acknowledgements

We would like to thank Alice Tillema, librarian at the Radboud University Medical Centre, for her help in defining and conducting the search applied in our review study.

Disclosure Statement

The authors have no conflicts of interest to report.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Agha, R. A., & Fowler, A. J. (2015). The Role and Validity of Surgical Simulation. *International Surgery, 100*, 350-357. <https://doi.org/10.9738/INTSURG-D-14-00004.1>
- Alken, A., Tan, E., Luursema, J. M., Fluit, C., & van Goor, H. (2013). Feedback Activities of Instructors during a Trauma Surgery Course. *The American Journal of Surgery, 206*, 599-604. <https://doi.org/10.1016/j.amjsurg.2013.03.011>
- Alken, A., Tan, E., Luursema, J. M., Fluit, C., & van Goor, H. (2015). Coaching during a Trauma Surgery Team Training: Perceptions versus Structured Observations. *The American Journal of Surgery, 209*, 163-169.

- <https://doi.org/10.1016/j.amjsurg.2014.02.009>
- Anderson, C. I., Gupta, R. N., Larson, J. R., Abubars, O. I., Kwiecien, A. J., Lake, A. D. et al. (2013). Impact of Objectively Assessing Surgeons' Teaching on Effective Perioperative Instructional Behaviors. *JAMA Surgery, 148*, 915-922. <https://doi.org/10.1001/jamasurg.2013.2144>
- Anton, N. E., & Stefanidis, D. (2016). *Should Surgeons Have Mental Skills Training?* Oxford: Oxford University Press. <https://doi.org/10.1093/ejcts/ezw071>
- Belfield, C., Thomas, H., Bullock, A., Eynon, R., & Wall, D. (2001). Measuring Effectiveness for Best Evidence Medical Education: A Discussion. *Medical Teacher, 23*, 164-170. <https://doi.org/10.1080/0142150020031084>
- Blom, E. M., Verdaasdonk, E. G., Stassen, L. P., Stassen, H. G., Wieringa, P. A., & Dankelman, J. (2007). Analysis of Verbal Communication during Teaching in the Operating Room and the Potentials for Surgical Training. *Surgical Endoscopy, 21*, 1560-1566. <https://doi.org/10.1007/s00464-006-9161-0>
- Chen, X. P., Williams, R. G., Sanfey, H. A., & Smink, D. S. (2015). A Taxonomy of Surgeons' Guiding Behaviors in the Operating Room. *The American Journal of Surgery, 209*, 15-20. <https://doi.org/10.1016/j.amjsurg.2014.07.018>
- Claridge, J. A., Calland, J. F., Chandrasekhara, V., Young, J. S., Sanfey, H., & Schirmer, B. D. (2003). Comparing Resident Measurements to Attending Surgeon Self-Perceptions of Surgical Educators. *The American Journal of Surgery, 185*, 323-327. [https://doi.org/10.1016/S0002-9610\(02\)01421-6](https://doi.org/10.1016/S0002-9610(02)01421-6)
- Cook, D. A., & Reed, D. A. (2015). Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. *Academic Medicine, 90*, 1067-1076. <https://doi.org/10.1097/ACM.0000000000000786>
- Dawe, S. R., Pena, G. N., Windsor, J. A., Broeders, J. A., Cregan, P. C., Hewett, P. J. et al. (2014). Systematic Review of Skills Transfer after Surgical Simulation-Based Training. *British Journal of Surgery, 101*, 1063-1076. <https://doi.org/10.1002/bjs.9482>
- De Win, G., Van Bruwaene, S., Kulkarni, J., Van Calster, B., Aggarwal, R., Allen, C. et al. (2016). An Evidence-Based Laparoscopic Simulation Curriculum Shortens the Clinical Learning Curve and Reduces Surgical Adverse Events. *Advances in Medical Education and Practice, 7*, 357-370. <https://doi.org/10.2147/AMEP.S102000>
- Duke, R. A., & Henninger, J. C. (1998). Effects of Verbal Corrections on Student Attitude and Performance. *Journal of Research in Music Education, 46*, 482-495. <https://doi.org/10.2307/3345345>
- Flinn, J. T., Miller, A., Pyatka, N., Brewer, J., Schneider, T., & Cao, C. G. (2016). The Effect of Stress on Learning in Surgical Skill Acquisition. *Medical Teacher, 38*, 897-903. <https://doi.org/10.3109/0142159X.2015.1114597>
- Gurusamy, K. S., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2009). Virtual Reality Training for Surgical Trainees in Laparoscopic Surgery. *Cochrane Database of Systematic Reviews, No. 1*, CD006575. <https://doi.org/10.1002/14651858.CD006575.pub2>
- Harrison, J. M., Fellingham, G. W., Buck, M. M., & Pellett, T. L. (1995). Effects of Practice and Command Styles on Rate of Change in Volleyball Performance and Self-Efficacy of High-, Medium-, and Low-Skilled Learners. *Journal of Teaching in Physical Education, 14*, 328-339. <https://doi.org/10.1123/jtpe.14.3.328>
- Hauge, L. S., Wanzek, J. A., & Godellas, C. (2001). The Reliability of an Instrument for Identifying and Quantifying Surgeons' Teaching in the Operating Room. *The American Journal of Surgery, 181*, 333-337. [https://doi.org/10.1016/S0002-9610\(01\)00577-3](https://doi.org/10.1016/S0002-9610(01)00577-3)

- Henninger, J. C., Flowers, P. J., & Council, K. H. (2006). Pedagogical Techniques and Student Outcomes in Applied Instrumental Lessons Taught by Experienced and Pre-Service American Music Teachers. *International Journal of Music Education, 24*, 71-84. <https://doi.org/10.1177/0255761406063108>
- Jensen, A. R., Wright, A. S., Kim, S., Horvath, K. D., & Calhoun, K. E. (2012). Educational Feedback in the Operating Room: A Gap between Resident and Faculty Perceptions. *The American Journal of Surgery, 204*, 248-255. <https://doi.org/10.1016/j.amjsurg.2011.08.019>
- Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under Stress: How Does It Work? *Trends in Cognitive Sciences, 10*, 152-158. <https://doi.org/10.1016/j.tics.2006.02.002>
- Kneebone, R. (2003). Simulation in Surgical Training: Educational Issues and Practical Implications. *Medical Education, 37*, 267-277. <https://doi.org/10.1046/j.1365-2923.2003.01440.x>
- Konge, L., & Lonn, L. (2016). Simulation-Based Training of Surgical Skills. *Perspectives on Medical Education, 5*, 3-4. <https://doi.org/10.1007/s40037-016-0251-y>
- Levinson, K. L., Barlin, J. N., Altman, K., & Satin, A. J. (2010). Disparity between Resident and Attending Physician Perceptions of Intraoperative Supervision and Education. *The Journal of Graduate Medical Education, 2*, 31-36. <https://doi.org/10.4300/JGME-D-09-00096.1>
- Li, M. M., & George, J. (2017). A Systematic Review of Low-Cost Laparoscopic Simulators. *Surgical Endoscopy, 31*, 38-48. <https://doi.org/10.1007/s00464-016-4953-3>
- McCaskie, A. W., Kenny, D. T., & Deshmukh, S. (2011). How Can Surgical Training Benefit from Theories of Skilled Motor Development, Musical Skill Acquisition and Performance Psychology? *The Medical Journal of Australia, 194*, 463-465.
- McKendy, K. M., Watanabe, Y., Lee, L., Bilgic, E., Enani, G., Feldman et al. (2017). Perioperative Feedback in Surgical Training: A Systematic Review. *The American Journal of Surgery, 214*, 117-126. <https://doi.org/10.1016/j.amjsurg.2016.12.014>
- McSparron, J. I., Ricotta, D. N., Moskowitz, A., Volpicelli, F. M., Roberts, D. H., Schwartzstein, R. M. et al. (2015). The PrOSTE: Identifying Key Components of Effective Procedural Teaching. *Annals of the American Thoracic Society, 12*, 230-234. <https://doi.org/10.1513/AnnalsATS.201406-237BC>
- Nicholls, D., Sweet, L., Muller, A., & Hyett, J. (2016). Teaching Psychomotor Skills in the Twenty-First Century: Revisiting and Reviewing Instructional Approaches through the Lens of Contemporary Literature. *Medical Teacher, 38*, 1056-1063. <https://doi.org/10.3109/0142159X.2016.1150984>
- Reznick, R. K., & MacRae, H. (2006). Teaching Surgical Skills—Changes in the Wind. *The New England Journal of Medicine, 355*, 2664-2669. <https://doi.org/10.1056/NEJMra054785>
- Ridder, J., McGaghie, W. C., Stokking, K. M., & Cate, O. T. (2015). Variables That Affect the Process and Outcome of Feedback, Relevant for Medical Training: A Meta-Review. *Medical Education, 49*, 658-673. <https://doi.org/10.1111/medu.12744>
- Roberts, N. K., Brenner, M. J., Williams, R. G., Kim, M. J., & Dunnington, G. L. (2012). Capturing the Teachable Moment: A Grounded Theory Study of Verbal Teaching Interactions in the Operating Room. *Surgery, 151*, 643-650. <https://doi.org/10.1016/j.surg.2011.12.011>
- Rogers, R. G. (2006). Mental Practice and Acquisition of Motor Skills: Examples from Sports Training and Surgical Education. *Obstetrics and Gynecology Clinics of North*

- America*, 33, 297-304. <https://doi.org/10.1016/j.ogc.2006.02.004>
<http://www.sciencedirect.com/science/article/pii/S0889854506000131>
- Rose, J. S., Waibel, B. H., & Schenarts, P. J. (2011). Disparity between Resident and Faculty Surgeons' Perceptions of Preoperative Preparation, Intraoperative Teaching, and Postoperative Feedback. *Journal of Surgical Education*, 68, 459-464. <https://doi.org/10.1016/j.jsurg.2011.04.003>
- Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress Effects on Memory: An Update and Integration. *Neuroscience & Biobehavioral Reviews*, 36, 1740-1749. <https://doi.org/10.1016/j.neubiorev.2011.07.002>
- Snyder, R. A., Tarpley, M. J., Tarpley, J. L., Davidson, M., Brophy, C., & Dattilo, J. B. (2012). Teaching in the Operating Room: Results of a National Survey. *Journal of Surgical Education*, 69, 643-649. <https://doi.org/10.1016/j.jsurg.2012.06.007>
- Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J., & Maddern, G. J. (2008). A Systematic Review of Skills Transfer after Surgical Simulation Training. *Annals of Surgery*, 248, 166-179. <https://doi.org/10.1097/SLA.0b013e318176bf24>
- 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, 291-300. <https://doi.org/10.1097/01.sla.0000200839.93965.26>
- Sutkin, G., Littleton, E. B., & Kanter, S. L. (2015). How Surgical Mentors Teach: A Classification of *in Vivo* Teaching Behaviors Part 1: Verbal Teaching Guidance. *Journal of Surgical Education*, 72, 243-250. <https://doi.org/10.1016/j.jsurg.2014.10.003>
- Timberlake, M. D., Mayo, H. G., Scott, L., Weis, J., & Gardner, A. K. (2017). What Do We Know about Intraoperative Teaching? A Systematic Review. *Annals of Surgery*, 266, 251-259. <https://doi.org/10.1097/SLA.0000000000002131>
- Vogel, S., & Schwabe, L. (2016). Learning and Memory under Stress: Implications for the Classroom. *NPJ Science of Learning*, 1, 16011. <https://doi.org/10.1038/npjscilearn.2016.11>
- White, C., Rodger, M. W. M., & Tang, T. (2016). Current Understanding of Learning Psychomotor Skills and the Impact on Teaching Laparoscopic Surgical Skills. *The Obstetrician & Gynaecologist*, 18, 53-63. <https://doi.org/10.1111/tog.12255>
- Wulf, G., Lauterbach, B., & Toole, T. (1999). The Learning Advantages of an External Focus of Attention in Golf. *Research Quarterly for Exercise and Sport*, 70, 120-126. <https://doi.org/10.1080/02701367.1999.10608029>
- Wulf, G., McConnel, N., Gartner, M., & Schwarz, A. (2002). Enhancing the Learning of Sport Skills through External-Focus Feedback. *Journal of Motor Behavior*, 34, 171-182. <https://doi.org/10.1080/00222890209601939>