BMJ Open

Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis

Christina Elizabeth Johnson [1,2] Mihiri P Weerasuria [3] Jennifer L Keating [4]

To cite: Johnson CE, Weerasuria MP, Keating JL. Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis. *BMJ Open* 2020;**10**:e030672. doi:10.1136/bmjopen-2019-030672

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-030672).

Received 28 March 2019
Revised 30 January 2020
Accepted 17 February 2020

For numbered affiliations see end of article.

Correspondence to
Dr Christina Elizabeth Johnson; christina.johnson@monashhealth.org

ABSTRACT

Objective Verbal face-to-face feedback on clinical task performance is a fundamental component of health professions education. Experts argue that feedback is critical for performance improvement, but the evidence is limited. The aim of this systematic review was to investigate the effect of face-to-face verbal feedback from a health professional, compared with alternative or no feedback, on the objective workplace task performance of another health professional.

Design Systematic review and meta-analysis.

Methods We searched the full holdings of Ovid MEDLINE, CENTRAL, Embase, CINAHL and PsycINFO up to 1 February 2019 and searched references of included studies. Two authors independently undertook study selection, data extraction and quality appraisal. Studies were included if they were randomised controlled trials investigating the effect of feedback, in which health professionals were randomised to individual verbal face-to-face feedback compared with no feedback or alternative feedback and available as full-text publications in English. The certainty of evidence was assessed using the Grading of Recommendations, Assessment, Development and Evaluations approach. For feedback compared with no feedback, outcome data from included studies were pooled using a random effects model.

Results In total, 26 trials met the inclusion criteria, involving 2307 participants. For the effect of verbal face-to-face feedback on performance compared with no feedback, when studies at high risk of bias were excluded, eight studies involving 392 health professionals were included in a meta-analysis: the standardised mean difference (SMD) was 0.7 (95% CI 0.37 to 1.03; p<0.001) in favour of feedback. The calculated SMD prediction interval was −0.06 to 1.46. For feedback compared with alternative feedback, studies could not be pooled due to substantial design and intervention heterogeneity. All included studies were summarised, and key factors likely to influence performance were identified including components within feedback interventions, instruction and practice opportunities.

Conclusions Verbal face-to-face feedback in the health professions may result in a moderate to large improvement in workplace task performance, compared with no feedback. However, the quality of evidence was low, primarily due to risk of bias and publication bias. Further research is needed. In particular, we found a lack of high-quality trials that clearly reported key components likely to influence performance.

Trial registration number CRD42017081796.

Strengths and limitations of this study

- This systematic review is the first to investigate the impact of face-to-face verbal feedback from a health professional, compared with alternative or no feedback, on the objective workplace task performance of another health professional.
- The meta-analysis of verbal feedback compared with no feedback is the first to provide an estimate of the likely benefit of verbal feedback on performance of a workplace task in the health professions.
- The quality of evidence was low, primarily due to risk of bias in study design or conduct and publication bias.

INTRODUCTION

Health professions education is embedded in clinical practice for both students and qualified staff as they continue learning and training.[1] Face-to-face verbal feedback focused on the performance of a clinical task involving an educator (senior clinician or peer) and a learner (any clinician) plays a crucial role in workplace learning, particularly within competency-based education and programmatic assessment models.[2–5]

Multiple reviews on feedback in health professional education have been published and include recommendations for

effective practice.[6–9] Feedback can occur in various forms, including verbal, written or automated (eg, from a simulator or within an online learning module). The unique potential benefits of face-to-face verbal feedback are the opportunities for: (1) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to coconstruct insights and strategies for improvement and (2) interpersonal connection, through which an educator can foster a learner's feelings of support, self-efficacy and motivation to improve, which are important catalysts in the learning process.[8 10–13]

There is widespread acceptance that feedback has an important role in maximising learning and achievement.[6 14–16] Ende[17] said, 'Without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all'. However, there is little evidence to support this view that feedback enhances health professionals' performance. Indeed, a recent scoping review on feedback identified the need for systematic reviews to support evidence-based recommendations.[7]

The current strongest evidence relates to two systematic reviews that investigated the impact of audit and feedback. In 2006, Veloski *et al*[18] published a BEME systematic review in which almost 75% of included studies reported that audit and feedback could improve an individual physician's clinical performance, particularly when sustained and from an authoritative source. Feedback was defined as 'summary information on clinical performance over a defined time period'. They included any empirical study (not just randomised controlled trials) and all types of physicians (most were primary care physicians). The majority of outcomes were clinical processes (such as test ordering), and the most common data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers *et al*[19] updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared with usual care. The review included various health professionals (predominantly doctors), the unit of allocation for interventions ranged from individuals to health services and the performance outcomes reported were clinical practice processes, such as the number or quality of prescriptions or tests. The authors argued that although the median risk difference (RD) in favour of feedback was small at 4.3% (IQR 0.5%–16%), the 3rd quartile at 16% indicated that audit and feedback interventions could be much more effective. Using multivariable meta-regression, they identified that the effectiveness of audit and feedback increased when the source was a senior colleague or supervisor (RD 11%), the format involved both written and verbal components (RD 8%), the frequency was at least monthly (RD 7%), the aim was to reduce specific behaviour (RD 6%) and it included both explicit measurable targets and a specific action plan involving advice on how to improve, compared with performance information alone (RD 5%). In addition, two other factors were associated with a higher likelihood of improvement: a lower baseline performance and the type of behaviour being targeted, for example, prescribing (possibly perceived as 'important' and 'straightforward') had better outcomes than improving diabetes management (more 'complex') or test ordering (possibly perceived as 'less important').

We found no systematic review that investigated the impact of verbal face-to-face feedback on a health professional's performance, the typical scenario in clinical practice.

Our research question was therefore:

'What is known about the effect of face-to-face verbal feedback from a health professional, compared with alternative or no feedback, on the objective performance of an observable workplace task by another health professional?'

The primary aim of the review addressed this question. Secondary aims were to summarise interventions and outcomes reported in included studies.

METHODS

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) statement.[20] The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO).

Eligibility criteria for considering studies for this review

We included randomised controlled trials in which individual health professionals were randomised to feedback, compared with no feedback or alternative feedback. Reports had to be available as English full-text publications.

We included studies in which participants were health professional students or graduates from the disciplines of medicine, dentistry, nursing and midwifery, allied health, psychology, pharmacy, medical radiation practice, optometry, osteopathy or chiropracty.

All studies had to include at least one intervention involving verbal face-to-face feedback generated by a health professional, based on the observable performance of a workplace task performed by another health professional. A broad definition of feedback was permitted with a minimum requirement that it included information regarding learner performance. Studies were excluded if feedback was predetermined or provided only by a simulated patient or machine. Audit and feedback studies, where feedback was based on aggregated quality performance indicators (such as numbers of tests ordered or degree of compliance with quality practice standards) were excluded, as this was deemed to be distinctly different from a workplace task, such as suturing, that could be observed, objectively assessed and targeted for improvement with feedback. Two comparisons were evaluated: (1) verbal face-to-face feedback compared with no feedback and (2) verbal face-to-face feedback compared with alternative feedback.

Performance following feedback interventions had to be objectively assessed. To isolate the effects of feedback,

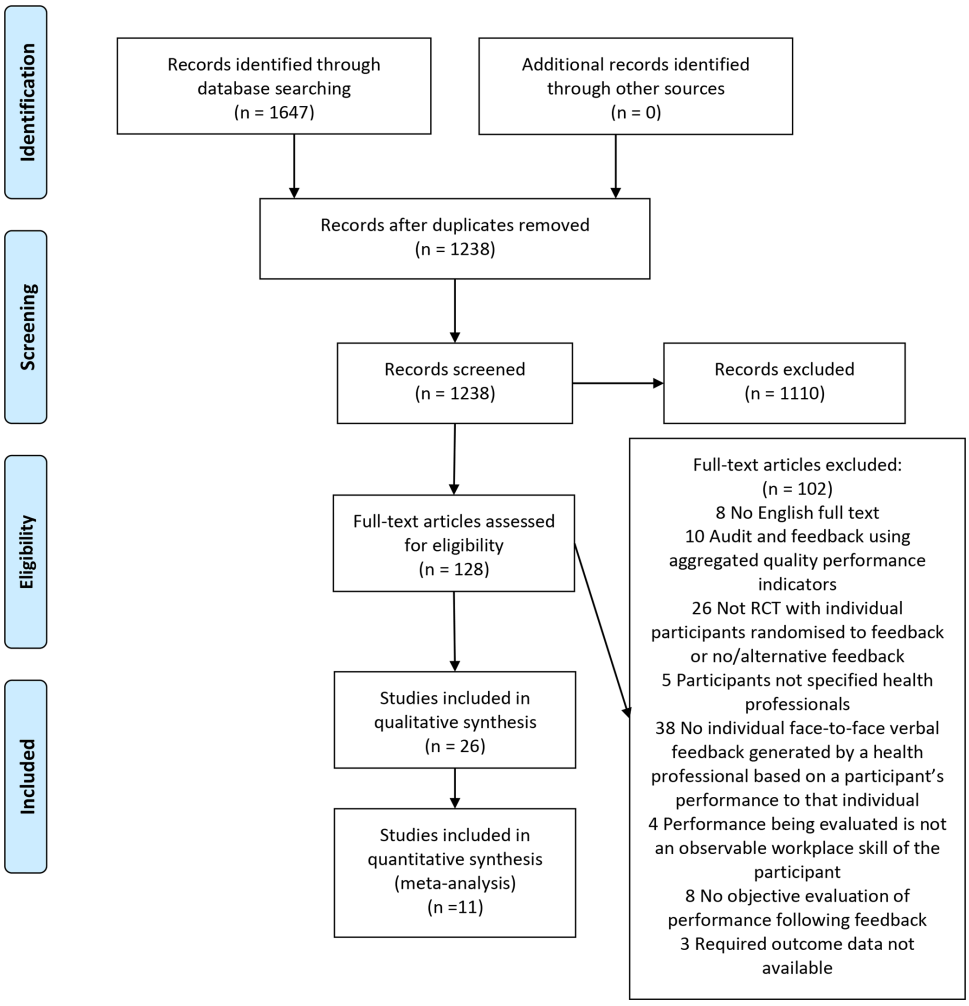

Figure 1 PRISMA flow diagram for systematic review of verbal face-to-face feedback compared with no or alternative feedback. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomised controlled trial.

other conditions had to be comparable for both groups. Studies were excluded if the report did not include point estimates of effects and measures of variability (or data from which these could be derived), unless these data could be obtained from the author.

Search methods for identification of studies

We developed the search strategy in collaboration with a senior medical librarian using MEDLINE subject headings. Key words were used, including synonyms, truncation, wildcard and proximity operators related to 'feedback' AND 'health professional' AND 'performance' AND 'randomised controlled trial' (see online supplementary appendix 1 for the full search strategy for Ovid MEDLINE). We translated this search strategy for other databases. The full holdings of Ovid MEDLINE (1946 to present with daily update), CENTRAL, Embase (1946 to present with daily update), CINAHL plus (1937 to present) and PsycINFO (1806 to present) were searched until 1 February 2019. We also searched the reference lists of systematic reviews and included studies.

Selection of studies

One review author (CEJ or MPW) screened titles to exclude clearly irrelevant reports. Two authors (CEJ and MPW) independently screened remaining abstracts to identify potential eligible studies, then independently assessed the full text. Decisions were compared using Covidence (online software designed by the Cochrane Collaboration to improve review efficiency via www.covidence.org), and disagreements were resolved through discussion, including a third review author (JLK).

Data extraction

One review author (CEJ) used a prepiloted standardised form to extract data from included studies, and another author (MPW or JLK) checked the data extracted were accurate. We resolved discrepancies through discussion. The following data were recorded: year of publication; study setting; funding sources; key details regarding participants, workplace task, feedback intervention and outcome measures; and information related to the risk of bias assessment. If data were missing, we contacted authors to request the information.

Table 1 Summary of available data on characteristics of trials included in the comparison of verbal face-to-face feedback (intervention) compared with no feedback (control: no feedback from any external source) on performance

Author Year Country	Task	Participants Health profession Experience % male	Teaching and practice Same for feedback intervention and control groups	Feedback intervention											Study outcomes* All effects are SMD (95% CI) and P valuefavours feedback
				Additional information	Source		Timing			Content					
					Subject expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info†	Performance video	Written performance info	
Ahlborg 2015[30] Sweden	Simulated laparoscopic O&G surgery using a VR simulator (salpingectomy).	Medical students. UGY5. 50% M.	Intervention duration: 1 session Case discussion+expert demonstration. 2× practice trials. Performance evaluation: end of session.	2× fb episodes. Fb given by expert (1) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus (2) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓			0.91 (–0.14 to 1.95) p=0.08
Bonrath 2015[34] Canada	GI surgery in routine clinical practice (jejuno-jejunostomy during laparoscopic bariatric surgery).	Doctors training in surgery. PGY3-5. 72% M.	Intervention duration: 2 months minimum. No teaching or practice in addition to routine clinical training. Performance evaluation: end of clinical attachment.	4 (approx.) × 25 min fb episodes. Fb given by expert using specific coaching model+video review of learner operating+video exemplars of good/poor technique. Effectiveness of strategies reviewed at subsequent session.	✓				✓	✓	✓		✓		1.62 (0.52 to 2.72) p=0.002
Boyle 2011[31] (*expert* fb) Ireland	Simulated endovascular surgery using a VR simulator (renal artery angioplasty +stent).	Doctors training in surgery. PGY4+.	Intervention duration: 1 session. Teaching+expert demonstration. 5× practice trials. Performance evaluation: end of session.	5× fb episodes. *Experts* provided 'whatever feedback they considered appropriate'+simulator output information.	✓			✓		?	?	✓			1.27 (–0.32 to 2.87) p=0.08
Boyle 2011[31] (*peer* fb)	Same as above.	Same as above.	Same as above.	5× fb episodes. *Peer* discussed simulator output, any task errors and teaching instructions given at start.		✓		✓		✓	✓	✓			0.81 (–0.66 to 2.29) p=0.24

Continued

Table 1 Continued

Author Year Country	Task	Participants Health profession Experience % male	Teaching and practice Same for feedback intervention and control groups	Feedback intervention											Study outcomes* All effects are SMD (95% CI) and P valuefavours feedback
				Additional information	Source		Timing			Content					
					Subject expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info†	Performance video	Written performance info	
Kroft 2017[35] Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy).	Doctors training in O&G. PGY2-6. 33% M.	Intervention duration: 1×15 min practice using laparoscopic salpingectomy module on VR surgical simulator within 1 hour of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1× fb episode from expert directly after VR simulator practice. Fb 'standardised and given in an evidence based fashion to optimise effectiveness' and included '3 constructive recommendations based on performance'.	✓			✓		✓	✓				0.85 (–0.35 to 2.06) p=0.14
O'Connor 2008[38] USA	Simulated surgical skill using a laparoscopic simulator (suturing and knot tying).	Medical students. UGY1–2. 44% M.	Intervention duration: 4 weeks. 2-hour instruction +practice suturing and knot tying until able to do it easily. Then instruction on laparoscopic surgery+expert demonstration video of task tying, followed by 30 min familiarisation with equipment. Practice: 1 hour daily, 6 days per week for 4 weeks. Performance evaluation: combined assessment of each attempt throughout intervention.	Expert fb provided 'continually on how to improve' during practice sessions+detailed explanations of simulator output information at the end of the session+given target performance goals.	✓		✓	✓		✓	✓	✓			0.40 (–1.25 to 2.04) p=0.58

Continued

Table 1 Continued

Author Year Country	Task	Participants Health profession Experience % male	Teaching and practice Same for feedback intervention and control groups	Feedback intervention											Study outcomes* All effects are SMD (95% CI) and P valuefavours feedback
				Additional information	Source		Timing			Content					
					Subject expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info†	Performance video	Written performance info	
Olms 2016[32] Germany	Simulated colour matching teeth.	Dental students. UGY3.	Intervention duration: 1 session. Study conducted during 10-week routine university module on matching tooth shades involving variety of teaching+practice opportunities. Performance evaluation: 2 weeks after intervention (within one university module).	1× expert fb session. Fb included correct response+explanation with expert demonstration if needed+written copy of evaluation. Expert trained in fb.	✓			✓		✓	✓			✓	2.09 (1.45 to 2.73) p<0.001
Pavo 2016[33] Austria	Simulated CPR.	Medical students. UGY3. 57% M.	Intervention duration: 1 session. Instruction on basic life support occurred previously, as part of university course. 1×2-hour additional training session: instructional video+training using modified Peyton four-step approach.‡ Brief practice (few minutes) in pairs using a manikin. Performance evaluation: end of session.	Fb during performance from peer performing ventilation to the student performing compressions (being assessed), at the start of each set of 30 chest compressions. Fb included information+corrective advice on compression rate and depth, hand position, decompression and hands-off time. Instructional video for intervention group had demonstrated this.		✓	✓			✓	✓				0.25 (−0.02 to 0.51) p=0.06

Continued

Table 1 Continued

Author Year Country	Task	Participants Health profession Experience % male	Teaching and practice Same for feedback intervention and control groups	Feedback intervention											Study outcomes* All effects are SMD (95% CI) and P valuefavours feedback
				Additional information	Source		Timing			Content					
					Subject expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info†	Performance video	Written performance info	
Skeff 1983[39] USA	Clinical teaching skills during ward round.	Physicians.	Intervention duration: 1 session in the middle of 4-week ward duty. At mid and end of ward duty: video of physician's teaching on ward rounds+rating of physician's teaching skills by medical students and junior medical staff on ward (video+ratings not shown to control group). Performance evaluation: 2 weeks later, at end of ward duty.	1× 60 min fb discussion with peer, including video review, trainee ratings and self-assessment to enable physician to identify strengths and devise solutions to problems.	✓				✓	✓	✓		✓	✓	0.56 (−0.15 to 1.27) p=0.12
Soucisse 2017[36] Canada	Simulated surgical procedure (bench-top intestinal anastomosis using cadaveric dog bowel).	Doctors training in surgery. PGY1-4.	Intervention duration: 1 session. Task instruction occurred previously (no teaching or practice within intervention). Baseline performance videoed. Performance evaluation: 3 weeks later (ongoing clinical work as a surgical resident).	1× 30 min expert fb sometime after baseline performance with video review of baseline performance+adapted coaching model including 2–3 suggestions for improvement+expert demonstration followed by learner demonstration of desired improvements, as required+action plan.	✓				✓	✓	✓		✓		0.3 (−0.44 to 1.05) p=0.42

Continued

Table 1 Continued

Author Year Country	Task	Participants Health profession Experience % male	Teaching and practice Same for feedback intervention and control groups	Feedback intervention											Study outcomes* All effects are SMD (95% CI) and P valuefavours feedback
				Additional information	Source		Timing			Content					
					Subject expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info†	Performance video	Written performance info	
Vafaei 2017[40] Iran	Chest ultrasound for trauma patients in emergency.	Doctors training in emergency. PGY4. 57% M.	Intervention duration: 1 session. Instruction for task occurred in previous training year (no teaching or practice within intervention). Baseline performance assessed. Performance evaluation: 2 months later (ongoing work as emergency resident).	1× 5 min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓					3.04 (1.95 to 4.13) p<0.001
Xeroulis 2007[37] (fb *after*) Canada	Simulated surgical skill using a bench-top model (suturing and knot tying).	Medical students. UGY1.	Intervention duration:1 session. Instructional video on task. Practice involved 19× trials in 1 hour. Performance evaluation: end of session.	Expert fb as needed (expert or learner initiated), *after* practice trials, involving constructive ways to improve+expert demonstration.	✓			✓		✓	✓				0.86 (–0.08 to 1.80) p=0.06
Xeroulis 2007[37] (fb *during*)	Same as above.	Same as above.	Same as above.	Same as above except expert fb *during* practice trials.	✓		✓			✓	✓				1.44 (0.43 to 2.46) p=0.004

*See 'Meta-analysis' section in Results for additional study details.
†Machine output information: simulator metrics (eg, procedural time or instrument path length) or CPR machine information (eg, compression rate and depth).
‡Peytons' 4-step model.[64]

CPR, cardiopulmonary resuscitation; fb, feedback; GI, gastrointestinal; M, male; O&G, obstetrics and gynaecology; PGY, postgraduate year; SMD, standardised mean difference; UGY, undergraduate year.

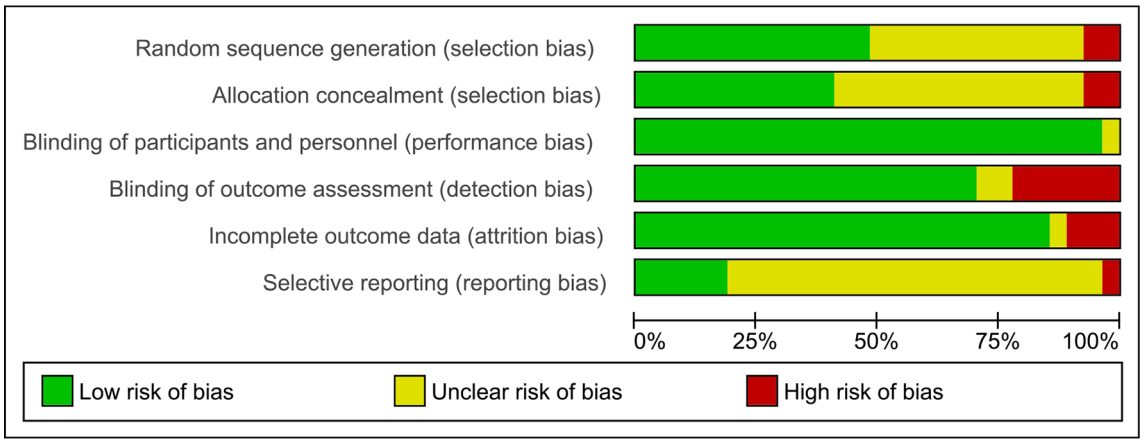

Figure 2 Risk of bias graph: review authors' judgements about each risk of bias itempresented as percentages across all included studies.

Assessment of risk of bias in included studies

The risk of bias was independently assessed by two authors (CEJ and JLK) for the selected performance outcome for individual studies, using Cochrane's 'risk of bias' tool (Chapter 8, Cochrane Handbook for Systematic Reviews of Interventions).[21] In particular, we used the following decision rules in assessing the risk of bias for specific individual domains. For 'participant and research team blinding': a participant receiving feedback or an educator giving feedback was deemed not to be blinded, even if they were deliberately not informed about the intervention or any differences between interventions. Nevertheless a 'low risk' rating was given if the outcome was not likely to be influenced by this lack of blinding, for example, if there were no changes to protocol or adherence that arose as a consequence of participant knowledge of group allocation.[22] For 'incomplete outcome data': to be rated as 'low risk', studies were required to include outcome data on at least 85% of the participants enrolled in each group (as per PEDRO guidelines),[23] and to provide participant numbers at the start and the number that dropped out during the study, from which group and the reasons.

The risk of bias was then summarised within each study across domains for the performance outcome, in accordance with the Cochrane 'risk of bias' assessment tool.

Measures of treatment effect

Outcomes from included studies were expressed using point estimates and measures of variability (eg, means (SD) or median (IQR). The effect was quantified using the standardised mean difference (SMD) to combine studies measuring the same outcome (task performance) using different measurement scales. When not reported, we estimated required data using available data or contacted study authors. If multiple outcomes were reported, we preferentially used the outcome that summarised multiple relevant task components, thereby providing a global, task-specific evaluation. If more than one reported outcome met this principle, we combined outcomes to provide a single metric using weighted averages of standardised scores.

We created and visually examined a funnel plot to explore reporting bias (Chapter 10, Cochrane Handbook).[24 25]

Data synthesis and assessment of heterogeneity

We pooled data from comparable studies for the comparison of feedback to no feedback on any measure of task performance and conducted analysis using random effects modelling in RevMan software (Review Manager Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). The result of the random effects meta-analysis was presented as the SMD of the treatment effect with 95% CI, as the average effect across multiple studies and its error estimates.

As a sensitivity analysis, we conducted a meta-analysis excluding studies with a high risk of bias. Using this pooled data, we calculated a prediction interval, which describes the range of likely results for new individual studies.[26]

We rated the overall certainty of evidence for the outcome using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach (Chapter 12, Cochrane Handbook[27 28] and GRADE guidelines),[29] which considers within-study risk of bias, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias. Two authors independently rated the certainty of the evidence and resolved disagreements by discussion. We presented a summary of the evidence in a 'Summary of Findings' table.

Patient and public involvement

There was no involvement of patients or the public in any part of this research.

RESULTS

Search results

The search yielded 1238 articles after 409 duplicates were removed. Based on title or abstract, we excluded 1110 articles. We assessed the remaining 128 full-text articles for eligibility and found 26 randomised controlled trials that met all inclusion criteria. See figure 1 for PRISMA study flow diagram.

Comparison 1: the effect of verbal face-to-face feedback, compared with no feedback, on performance

Included studies

Table 1 presents the characteristics of included studies in this comparison. Eleven randomised controlled trials investigated the effect of verbal face-to-face feedback compared with no feedback on the objective evaluation of a workplace task. Seven (64%) reports were published in the last 5 years since 2014. The studies were conducted in Europe (4),[30–33] Canada (4),[34–37] the USA (2)[38 39] and Asia (1).[40]

There were 488 participants, including 196/366 (53.6%) men from seven studies that reported gender data.[30 32–35 38 40] Participants included 290 (60%) medical students in four studies,[30 33 37 38] 60 (12%) dental students in one study[32] and 138 (28%) doctors in six studies.[31 34–36 39 40] The workplace tasks involved a discrete task such as surgical procedures, cardiopulmonary resuscitation (CPR) or teaching skills, which occurred in clinical practice in four studies[34 35 39 40] and a simulation environment in seven studies (7/11, 64%).[30–33 36–38] Differences in feedback interventions between included studies involved feedback source (expert or peer), timing (during task performance, directly afterwards or delayed), content (evaluative information only or additional corrective advice, performance video, simulator information or written report) and number of feedback episodes. In addition, there was variation between studies in provision of instruction and expert demonstration of the task, opportunities for practice and duration of feedback intervention (see 'Included studies' section in the supplementary material for more details, online supplementary appendix 2).

Risk of bias

The risk of bias graph is presented in figure 2, and the risk of bias summary is presented in figure 3. In summarising the risk of bias across domains within each study, two studies were rated 'low risk',[34 36] six studies were rated 'unclear'[30 31 33 35 37 39] and three studies were 'high risk'[32 38 40] (see 'Risk of bias' section in the supplementary material for more details, online supplementary appendix 2).

Reporting bias

The funnel plots are presented in figure 4: for all included studies (figure 4A) and after excluding studies at high risk of bias (figure 4B). Both funnel plots are asymmetrical, with a paucity of small studies with negative effect sizes that are less likely to be published, indicating some potential for publication bias.

Meta-analysis

A meta-analysis of the impact of verbal face-to-face feedback compared with no feedback on performance included 13 comparisons from the 11 studies, involving 488 participants. Two studies reported data that each enabled two comparisons: in one study, feedback

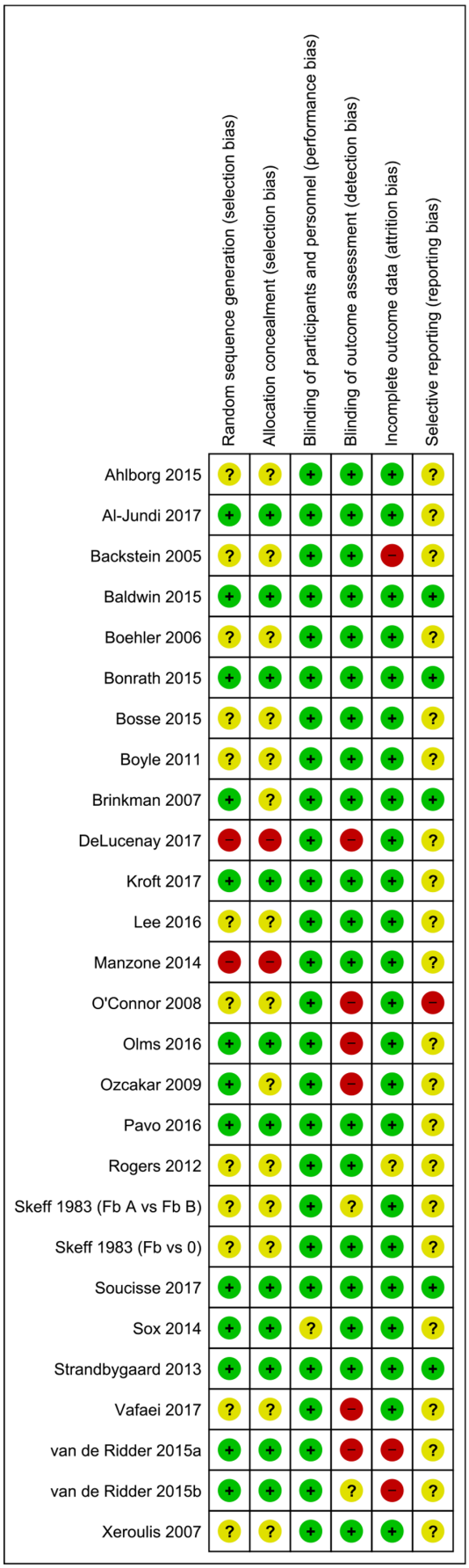

Figure 3 Risk of bias summary: review authors' judgements about each risk of bias item foreach included study.

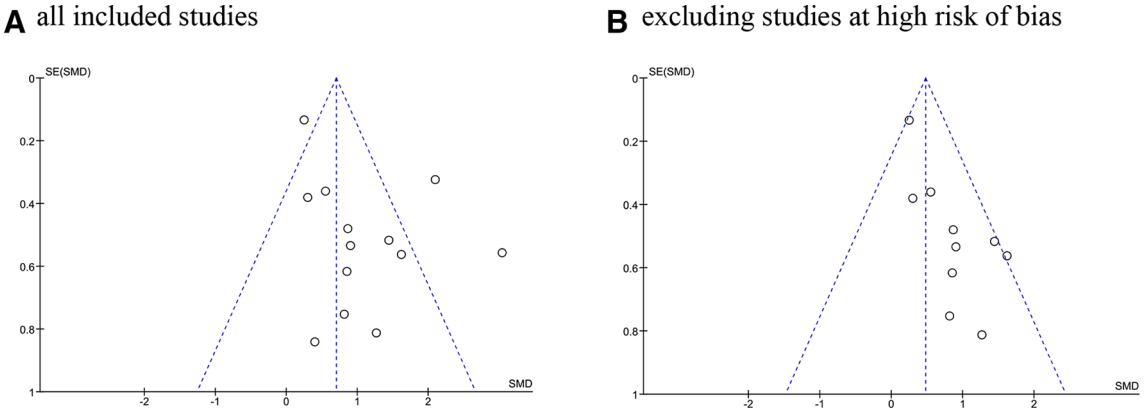

Figure 4 Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared with no feedback, on performance. Meta-analysis calcuated using a fixed effects model. The dotted vertical line represents the overall effect estimate and the dotted slanted lines represent the 95% CI. SMD, standardised mean difference.

provided *during* practice in one group and *directly after* practice in another were compared with no feedback[37]; in another study, feedback provided by an *expert* in one group and by a *peer* in another group[31] were compared with no feedback. In the meta-analysis, numbers for the control group for these studies were halved to retain sample independence.[27]

The meta-analysis of the effect of verbal face-to-face feedback compared with no feedback on workplace task performance found an SMD of 1.09 (95% CI 0.59 to 1.59; p<0. 001) using a random effects model. The forest plot is presented in figure 5A.

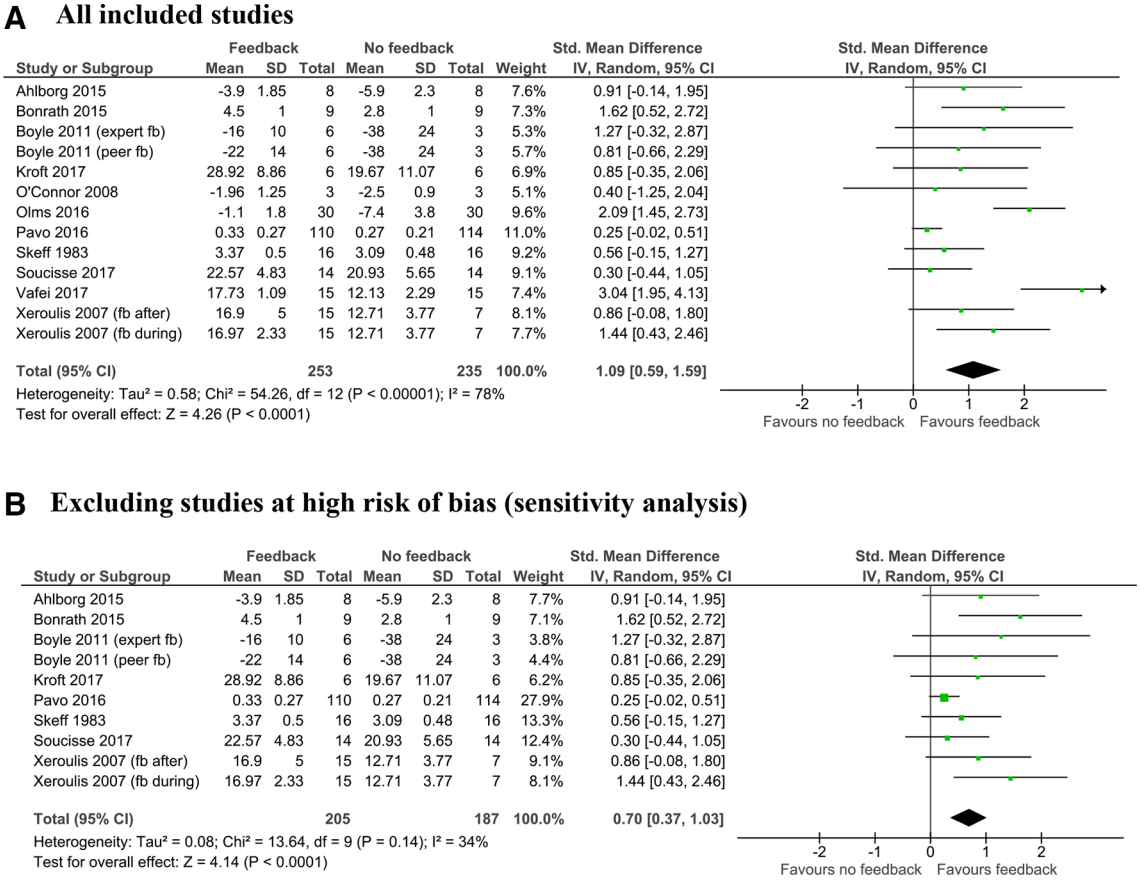

Figure 5 Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared with no feedback, on performance. Ahlborg 2015: mean and SD read from graph. Boyle 2011: mean and SD read from graph. Bonrath 2015: combined outcome calculated. Pavo 2016: median taken as best estimate of mean and calculated SD from IQR. Xeroulis 2007: SD estimated from 95% CI. SMD, standardised mean difference.

Verbal face-to-face feedback compared to no feedback for workplace task performance				
Patient or population: health professionals **Setting:** authentic or simulated clinical environment **Intervention:** verbal face-to-face feedback **Comparison:** no feedback				
	Standardised mean difference and 95% CI			
Outcomes	**With feedback**	**Participants**	**Certainty of evidence (GRADE)**	**Comments**
Objective assessment of observed performance	The mean score in the intervention group was 0.7 standard deviations (0.37 to 1.03) higher than mean scores for the control group	Number of participants 392 (8 studies)	⊕⊕⊖⊖[a,b] **low** Due to risk of bias and publication bias	Face-to-face feedback may result in a moderate to large improvement in workplace task performance
CI = Confidence interval; SMD= standardised mean difference				
[a] High risk of bias due to lack of allocation concealment and prior published protocols to preclude selective reporting of outcomes. [b]High probability of publication bias				

Figure 6 Summary of findings table for the effect of verbal face-to-face feedback, compared with no feedback, on performance,excluding studies with a high risk of bias.

Sensitivity analysis

As a sensitivity analysis, we repeated the random effects meta-analysis after excluding studies with a high risk of bias. Eight studies (8/11, 73%) were included that involved 392 health professional learners across 10 comparisons.[30 31 33–37 39] The SMD was 0.7 (95% CI 0.37 to 1.03; p<0.001). The forest plot is presented in figure 5B. The prediction interval was −0.06 to 1.46.

We judged that the certainty of the evidence was low, using the GRADE approach. We downgraded the overall rating from high to low, in view of a serious risk of bias (in particular, due to a lack of concealment and potential for selective reporting of outcomes) and publication bias[41] (see 'Certainty of evidence' section in the supplementary material for more details, online supplementary appendix 2). Figure 6 displays the summary of findings table.

Comparison 2: the effect of verbal face-to-face feedback, compared with alternative feedback, on performance

Included studies

Table 2 presents the characteristics of included studies in the comparison of verbal face-to-face feedback compared with alternative feedback. Twenty studies (22 comparisons) were included in this analysis and involved verbal, face-to-face feedback compared with alternative feedback. Nine studies (9/20, 45%) were published in the last 5 years since 2014. The studies were conducted in Europe (8/20, 40%), USA (7/20, 35%), Canada (4/20, 20%) and Asia (1/20, 5%).

There were 1974 participants, including 660/1463 (45%) men from 13 studies that reported gender data.[33 38 42–52] Included studies involved students (medical, mixed health professions and pharmacy) (1869, 95%) in 16 studies,[33 37 38 42–45 47–55] and doctors (105, 5%) in four studies.[31 39 46 56] All studies included assessment of a discrete task except two studies that involved longitudinal evaluations.[39 46] Three studies evaluated performance in a clinical practice setting (involving teaching skills,[39] professional and communication skills[46] and oral case presentations),[55] and the remaining 17 assessed performance in a simulated environment (surgical procedures, nasogastric tube insertion, intubation, hearing test, pharmacy consultation or CPR)[31 33 37 38 42–45 47–54 56] (see 'Included studies' section in the supplementary material for more details, online supplementary appendix 3).

Risk of bias

In summarising the risk of bias across domains within each study, two studies were rated as low risk,[43 50] seven studies were rated as 'high risk'[38 47 51 52 54 56] and the remaining studies were rated as 'unclear' (see the risk of bias summary in figure 3) (see 'Risk of bias' section in the supplementary material for more details, online supplementary appendix 3).

Effect of interventions

Figure 7 presents the forest plot and SMD. One additional study[43] that reported categorical data is not included in the forest plot. It compared a learning conversation (315 participants, pass rate 80.9%) to a feedback sandwich (325 participants, pass rate 77.2%) resulting in an OR of 1.25 (95% CI 0.85 to 1.84) that favoured the learning conversation. The feedback comparisons were markedly diverse, so we did not pool outcomes in meta-analysis.

DISCUSSION

Comparison 1: the effect of verbal face-to-face feedback, compared with no feedback, on performance

Our meta-analysis found that verbal face-to-face feedback may result in a moderate to large improvement in health professionals' performance compared with no feedback, with SMD 0.7 (95% CI 0.37 to 1.03; p<0.001)

Table 2 Summary of available data on characteristics of trials comparing the effect of verbal face-to-face fb (intervention A), to alternative fb (intervention B), on performance

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
Al-Jundi† 2017[42] England	Simulated surgical skill using bench top model ('skin' suturing with a latex pad).	Medical students. UGY5. 65% M.	Intervention duration: 1 session. Video instruction on surgical skill. 1×10 min for baseline performance. Performance evaluation: 2 days later.	Immediate face-to-face+written expert feedback. 1× expert fb. Expert observed baseline performance and rated it using task-specific checklist. Learner completed written self-assessment using same check list. Fb directly after performance, by expert with medical education qualification. Fb included verbal fb based on assessment checklist, 'directive and specific'+demonstration of skill, as required. Learner given copy of assessment+written feedback forms.	Delayed written expert fb via email. 1× written expert fb via email same day as baseline performance. Expert watched video of baseline performance, rated it using task-specific checklist and wrote fb comments aligned with assessment checklist, including suggestions for improvement, so fb was 'directive and specific'. Both assessment and written feedback forms emailed to learner.	−1.53 (−2.28 to −0.79) p<0.001 favours fb intervention B.
Backstein 2005[56] Canada	Simulated surgical procedure using a bench top model (vascular anastomosis).	Doctors in surgical training. PGY1.	Intervention duration: 4-week lecture on surgical procedure. 3×2-hour weekly practice sessions with expert fb as needed. Expert vascular surgeons undertook fb training, based on evaluation checklist and given in a similar manner. Performance evaluation: in week 4.	Review of performance video with expert fb +practice sessions with expert fb available. 3× weekly videotaping of surgical procedure, with expert fb available during task, followed by up to 15 min review of video with expert fb.	Practice sessions with expert fb available.	0.86 (0.05 to 1.67) p=0.03
Baldwin 2015[43] England	Simulated BLS	Health professional students. Medical (58%), physio (12%), pharmacy (10%), nursing (10%), dentistry (10%). UGY1. 33% M.	Intervention duration: 4 weeks. Instruction and practice with manikin 3×2.5 hours weekly. Fb provided directly after performance by senior peer instructor. Instructor accredited in BLS+trained to provide fb. Fb provider compliance monitored. Performance evaluation: in week 4.	'Learning conversation' model. Fb focused on learner's perspective: started with learner self-assessment, then explored issues and ideas raised by learner with group using advocacy inquiry format‡ with final summary.	'Feedback sandwich model'. Fb involved a point for improvement in between 2 points of praise.	OR 1.25§ (0.85 to 1.84) p=0.25
Boehler 2006[44] USA	Simulated surgical skill using a bench top model (tying a two-handed square knot)	Medical students. UGY2-3. 52% M.	Intervention duration: 1 session. Instruction in knot tying from surgeon. 1× baseline performance. Performance evaluation: end of session.	Expert feedback. 1× episode of fb from expert surgeon, directly after performance, describing 1–2 specific ways to improve performance.	Compliment. 1× prescripted general compliment, for example, 'great job!'.	0.98 (0.25 to 1.71) p=0.01

Continued

Table 2 Continued

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
Bosse 2015[45] Germany	Simulated nasogastric tube insertion (NGTI) into manikin.	Medical students. UGY1-2. 51% M.	Intervention duration: 1 session. NGTI training using case study role play and four-step procedural training method.¶ 6× practice trials. Fb 'positively worded', focused on effect of actions, given directly after performance by senior peer instructors, trained in procedure and fb. Performance evaluation: end of session.	High frequency fb. 6× episodes of fb, given after each practice trial.	Low frequency practice. 2× episodes of fb, given after first and last practice trial.	0.81 (0.21 to 1.40) p=0.01
Boyle 2011[31] Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty+stent).	Doctors training in surgery PGY4+.	Intervention duration: 1 session. Teaching and expert demonstration. Fb providers had simulator training. 5× practice trials (each maximum 40 min). Performance evaluation: end of session.	Expert fb. 5× fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	Peer fb. 5× fb episodes. Peer discussed simulator output, any task errors and the teaching instructions given at start of session.	0.46 (−0.70 to 1.61) p=0.41
Brinkman 2007[46] USA	Professional and communication skills during routine clinical practice on a paediatric ward.	Doctors training in paediatrics. PGY1. 34% M.	Intervention duration: 1 session. No teaching or practice within intervention. Routine fb as part of clinical training: monthly written evaluations from paediatricians on ward duty. Performance ratings obtained from nurses and patients at start and end of doctors' rotation. Performance evaluation: 5 months after start of clinical attachment.	Coaching session+routine fb as part of clinical training 1×30 min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses and parents. Used a coaching model focused on assisting learner to understand information, design goals and improvement strategies. Fb given by paediatricians trained in coaching model.	Routine fb as part of clinical training. Performance ratings from nurses and patients not seen.	2.70 (1.75 to 3.64) p<0.001

Continued

Table 2 Continued

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
DeLucenay 2017[47] USA	Simulated pharmacist patient consultation (identifying prescription errors and communication skills).	Pharmacy students. UGY3.	Intervention duration: 1 semester. Study conducted during usual university module on medication counselling involving 15 min SP consultations, each on a different topic. Directly after each one, SP provided 5 min fb on communication skills. Performance evaluation: last four SP consultations.	Immediate face-to-face fb. 4× expert fb directly after SP consultation and SP fb, based on expert's direct observation of SP consultation (unseen by participants). Fb included performance grade, performance and topic discussion with suggested improvements.	Delayed written fb. 4× videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.	0.30 (−0.02 to 0.62) p=0.07
Lee 2016[48] Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis).	Medical students. UGY3-4. 78% M.	Intervention duration: 3 weeks. Instruction and expert demonstration of procedure, followed by 3× weekly 30 min practice sessions. Performance evaluation: end of third session.	Early fb. 1× 10–15 min expert fb directly after first practice attempt, focused on assessment domains.	Late fb. Same as early fb but at end of second practice session.	1.3 (0.26 to 2.34) p=0.01
Manzone 2014[54] (verbal comment focused on performance vs verbal comment +comparison with training levels) Canada	Simulated intubation using manikin.	Medical students. UGY1-2.	Intervention duration: 1 session. Instructional video on intubation. 1–1.5 hour practice with manikin, with learner in four different positions (5× practice trials in each position). 10× fb by expert, given directly after practice trials in two positions (2×5). Fb only provided performance evaluation, with no advice on how to improve. Performance evaluation: end of session.	Performance comment focused on task. Fb involved evaluative performance comment, focused on any two aspects of performance (either done correctly or not), for example, 'improper use of the laryngoscope'. +individual's progress on task.	Performance comment compared with others (different training levels). Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standards at different training levels, for example, 'your performance was at the level of a resident'.	−0.93 (−1.89 to 0.03) p=0.05 favours fb intervention B.
Manzone 2014[54] (verbal comment on performance vs numerical rating, focused on individual progress)	As above.	As above.	As above.	Performance comment focused on task. As above.	Numerical performance outcome, focused on task progress. Provided with numerical performance information (performance time and number of hand movements). Plotted on graph to focus on own progress.	−0.37 (−1.26 to 0.51) p=0.39 favours fb intervention B.

Continued

Table 2 Continued

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
Manzone 2014[54] (verbal comment focused on performance vs numerical fb+comparison)	As above.	As above.	As above.	Performance comment focused on task. As above.	Numerical performance outcome, compared with others (scores at different training levels). Provided with numerical performance information (performance time and number of hand movements), accompanied by a list of scores across different training levels from medial student to specialist.	–2.87 (–4.20 to –1.55) p<0.001 favours fb intervention B.
O'Connor 2008[38] USA	Simulated surgical skill using a laparoscopic simulator (suturing and knot tying).	Medical students. UGY1-2. 44% M.	Intervention duration: 4 weeks. First session: 2-hour instruction and practice suturing and tying knots 'until able to do it easily'. Second session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30 min familiarisation with equipment. Practice: 1 hour daily, 6 days per week for 4 weeks. Simulator output information available at the end of each practice session: task completion time, smoothness of tool manipulation and path length of tool.	Expert fb during practice+simulator output information with expert discussion. Fb by surgical expert occurred continually throughout practice sessions. Expert observed participants closely, corrected mistakes early and provided instructions on how to improve + simulator output information with expert explanation of this information and given target goals.	Simulator output information.	0.51 (–1.16 to 2.19) p=0.48
Ozcakar 2009[49] Turkey	Simulated patient consultation with a simulated patient (communication and history-taking skills).	Medical students. UGY2. 62% M.	Intervention duration: 2 weeks. Study conducted during routine university module on clinical skills training. Evaluation: 2 weeks after intervention following clinical skills lectures+practice with video recording.	Video review with expert+expert fb. 1× videotaping of SP consultation. Directly afterwards, review video with expert plus fb.	Expert fb. 1× expert fb directly after SP consultation.	0.32 (–0.23 to 0.87) p=0.24
Pavo 2016[33] Austria	Simulated CPR.	Medical students. UGY3. 57% M.	Intervention duration: 1 session. Instruction on basic life support occurred previously, as part of university course. 1× 2-hour additional session including training using modified Peyton 4-step approach[e] and practice on a manikin. Performance evaluation: CPR skills at end of session.	Verbal fb from peer during CPR. Fb during performance from peer performing ventilation to the student performing compressions (being assessed), at the start of each set of 30 chest compressions. Fb included information+corrective advice on compression rate and depth, hand position, decompression and hands-off time. Brief practice by pair of participants with a manikin, until felt confident.	Machine output during CPR. CPR machine showed real-time visual display (numbers and graphs) of compression rate and depth plus automated audio advice to correct any deviations during CPR.	–0.09 (–0.36 to 0.18) p=0.53 favours fb intervention B

Continued

Table 2 Continued

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
Rogers 2012[53] USA	Simulated surgical skill (tying a single two-handed square knot).	Medical students. 'Surgical clerkship year'.	Intervention duration: 1 session. Training in knot tying. 2× practice trials (one before and one after training). Performance evaluation: end of session.	Expert fb. 1× fb from expert, with specific information on improving subsequent performance, directly after performance.	Compliment. 1× general compliment from expert, instead of fb.	1.69 (1.06 to 2.32) p<0.001
Skeff 1983[39] USA	Clinical teaching skills during ward round in routine clinical practice.	Attending physicians.	Intervention duration: 1 month. Performance evaluation: medical students and junior medical staff (trainees) on ward rated physicians' teaching skills during ward rounds, at the middle and end of 1-month term.	Expert peer fb. 1× 1-hour session midterm with expert peer, including review of videos of physician's teaching on ward rounds, trainees' evaluations and self-assessment of teaching skills. Fb discussion aimed at helping physician clarify strong teaching skills and devise solutions for teaching problems.	Written fb. Received written summary of trainees' evaluation of physician's teaching skills.	–0.36 (–1.06 to 0.34) p=0.30 favours fb intervention B
Sox 2014[55] USA	Case presentation during student clinical attachment in paediatrics.	Medical students. UGY3.	Intervention duration: paediatric clerkship. Week 1: lecture on important aspects of case presentations. Week 2: present case to small group with doctor in paediatric unit who was trained in evaluation. Performance evaluation: end of clerkship.	Detailed evaluation form. 1× constructive expert fb, directly after performance informed by 18-item evaluation form. Learner saw 18-item evaluation form but not given a copy.	Simple evaluation form. 1× constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw one item evaluation form but not given a copy.	0.15 (–0.07 to 0.37) p=0.17
Strandbygaard 2013[50] Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extrauterine pregnancy).	Medical students. UGY 4–6. 44% M.	Intervention duration: 2 months. 1× session with instruction expert demonstrations on operational technique, how to use simulator and interpret simulator output information. Simulator output information available after every practice: procedural time+performance score derived from multiple task performance criteria. Participants instructed to practice until achieved predefined expert proficiency level; could practice daily (max 3 hours) for up to 2 months.	Standardised expert fb with later, additional expert fb if requested by learner+simulator performance score. 1–3× 10–12 min episodes of expert fb involving information on how to perform task components correctly. First fb episode provided after first practice trial; learner could ask for up to two additional fb episodes (optional) involving same standardised advice+simulator performance score.	Simulator performance score.	1.31 (0.86 to 1.77) p<0.001

Continued

Table 2 Continued

Article First author Year Country	Task	Participants: health profession Experience Gender: % Men	Common to interventions A+B	Intervention A All included verbal face-to-face fb to an individual health professional	Intervention B	Study outcomes* Unless otherwise stated, effects are SMD (95% CI) and P value favours feedback intervention A
Van de Ridder 2015a[52] (Advances in Health Science Education) Netherlands	Simulated hearing test with a simulated patient (W&R test).	Medical students. UGY1. 35% M.	Intervention duration: 1× session. Instructional video of task. 1× baseline performance. Fb from senior medical student with acting experience and trained to act as a physician familiar with W&R test. Fb provider trained to give corrective information, cast in positive or negative tone according to study group allocation. Performance evaluation: end of session.	Positively framed fb. 1× fb directly after baseline performance. Fb comment started with global praise followed by the most suitable suggestion for improvement, selected from a list of four the most common task errors (eg, 'You did this well; a tip is …').	Negatively framed fb. 1× episode fb directly after practice trial. Fb comment started with global criticism followed by most appropriate directive advice for improvement, selected from list of four the most common task errors. (eg, 'You did not do this correctly; you should change.').	0.41 (−0.06 to 0.87) p=0.08
Van de Ridder 2015b[52] (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (W&R test).	Medical students. UGY1. 31% M.	Intervention duration: 1× session. Instructional video of task. 1× baseline performance. All fb providers trained for 1 hour on W&R test and giving fb according to protocol. Fb monitored to ensure it was given as per protocol. Performance evaluation: end of session.	High credibility fb provider. 1× fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag and introduced as professor ENT, wearing a white coat).	Low credibility fb provider. 1× fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female and informally dressed).	−0.23 (−0.71 to 0.26) p=0.36
Xeroulis 2007[37] Canada	Simulated surgical skill using a bench-top model (suturing and knot tying).	Medical students UGY1.	Intervention duration: 1 session. Instructional video on task. Practice involved 19 × trials in 1 hour. Fb involved constructive ways to improve+expert demonstration. Performance evaluation: end of session.	Expert fb during practice. Expert fb as needed (expert or learner initiated), *during* practice trials.	Expert fb directly after practice. Same as 'during practice' except fb *after* practice trials.	0.02 (−0.70 to 0.73) p=0.96 favours fb intervention B

*See figure 7 forest plot for additional study details.
†Additional data obtained from authors, enabling calculation of mean, SD and % men.
‡Advocacy inquiry approach.[73]
§Categorical data only available (see text in Results for more details).
¶Peyton's 4 steps.[64]
BLS, basic life support; CPR, cardiopulmonary resuscitation; ENT, ear, nose and throat specialist; fb, feedback; GRS, global rating scale; Max, maximum; NG, nasogastric; PGY, postgraduate year (referring to postqualification year); physio, physiotherapy; SMD, standardised mean difference; SP, simulated patients; UGY, undergraduate year (referring to university year level); VR, virtual reality; W&R, Weber & Rinne test.

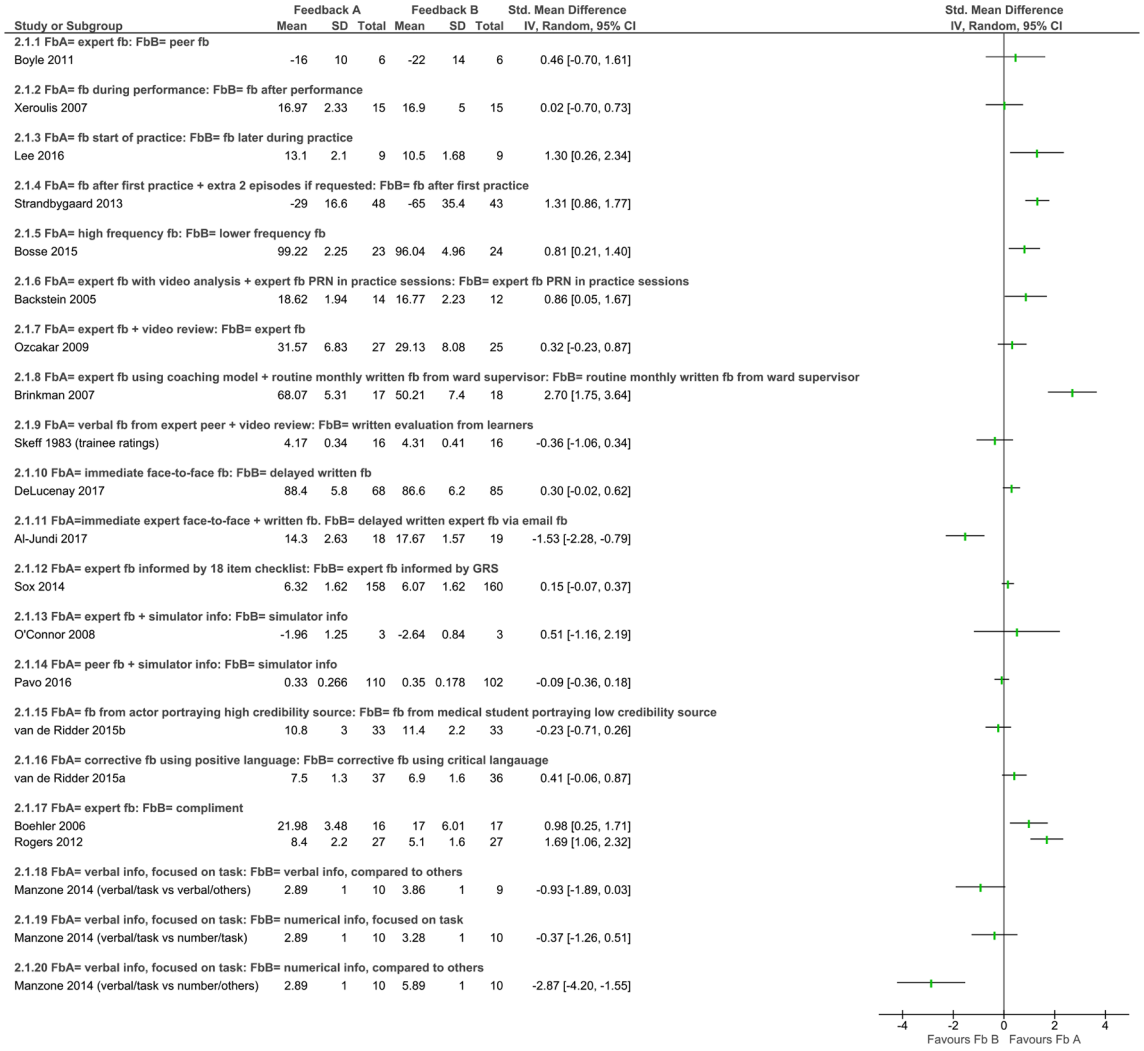

Figure 7 Forest plot for the effect of verbal face-to-face feedback (feedback A), compared with alternative feedback (feedback B), on performance. Baldwin 2015: categorical data not included in this figure; see text in Results. Al-Jundi 2017: additional information (data to calculate mean and SD for each group) from author. Boehler 2006: additional information (number of participants in each group and SD) from author. Lee 2016: calculated SD from SE. Manzone 2014: calculated standardised score to combine outcome of supine and normal positions. Pavo 2016: median taken as best estimate of mean. Rogers 2012: additional information (standard deviation) from author. Sox 2014: SD derived from reported t, p and mean values. Assumption that SDs were equivalent for intervention and controls. Strandbygaard 2013: SE derived from 95% CI. fb, feedback; GRS, global rating scale; info, information; PRN, 'as required'.

from eight studies involving 392 health professionals, after excluding studies at high risk of bias. However, the quality of evidence was low, primarily due to risk of bias and publication bias. To our knowledge, this is the first report to provide some substantiation for the widely held view that feedback enhances health professionals' performance and to estimate the benefit (see 'Discussion' section in the supplementary material for more details, online supplementary appendix 4).

The consistent positive effects across all included studies, with substantially overlapping confidence intervals, supports the likelihood that verbal face-to-face feedback enhances performance in the health professions. Our pooled effect size was moderate to large at 0.7.[57] The calculated prediction interval for the comparison of verbal face-to-face feedback to no feedback (excluding studies with a high risk of bias) was −0.06 to 1.45. This indicates a wide range in the likely feedback effect for any individual study, from a very small detrimental effect to a very large beneficial effect on performance. These results align with previous meta-analyses within health and other professions that have reported beneficial but variable effect sizes with different feedback interventions.[14 18 19] For example, within the health professions, Ivers *et al*[19] reported that 0.5% to 16% more participants followed desired practice when involved in an audit and feedback intervention. In comparison, a meta-analysis by Kluger and DeNisi,[58] which analysed any type of feedback across any discipline, compared with no feedback, reported a pooled SMD of 0.4; notably one-third of included studies reported a detrimental impact.

One possible explanation for this variability is that some constituents within a feedback intervention are more effective than others. When specific feedback elements were isolated, the largest beneficial effects of feedback reported in Kluger and DeNisi's meta-analysis were: (1) effect size 0.55 when feedback included information on any changes since the previous attempt, (3) effect size 0.51 when a specific and challenging goal was set, (3) effect size 0.47 when feedback posed little threat to self-esteem and (4) effect size 0.43 when feedback included information on the correct outcome.[58]

Comparison 2: the effect of verbal face-to-face feedback, compared with alternative feedback, on performance

For the second comparison of the effect of verbal face-to-face feedback compared with alternative feedback on performance, there was a diverse range in the alternative feedback interventions, which precluded meta-analysis. Where individual studies tested the relative impact of different feedback interventions, there was greater performance improvement seen with the following strategies: additional expert coaching sessions compared with routine monthly written feedback from supervisors[46]; expert feedback early in a practice period compared with later[48]; additional episodes of feedback from experts[45 50]; additional episodes of feedback involving expert video analysis[56]; and expert feedback compared with compliments.[44 53]

Influences on performance due to variations in the constituents of feedback interventions

The studies assembled in this review illustrate the wide variety of possible constituents within feedback interventions and the potential influence on performance. Within verbal face-to-face feedback interventions, there were important differences between included studies in feedback content, source and timing. Previous studies have noted potential beneficial effects attributable to feedback that contains information to clarify the goal[10 19 31 58] is delivered by educators with perceived credibility[18 19 59–63] and strategic use of both early and delayed feedback[48 59] (see 'Discussion' section in the supplementary material for more details, online supplementary appendix 4).

Influences on performance due to factors beyond feedback

Performance improvement is not solely related to feedback. In our review, other important factors influencing performance, such as instruction and practice opportunities, also varied between studies. These included teaching and expert demonstration,[37 48 59 64–69] learners' background, task complexity and practice opportunities[10 15 70 71] (see 'Discussion' section in the supplementary material for more details, online supplementary appendix 4).

Review limitations

The review has a number of limitations. Despite our attempts to be thorough, we may have missed studies that should have been included. As a number of studies did not report data that would allow easy pooling of data, we either calculated an estimate from available data (including reading off graphs) or excluded the study. Most included studies were conducted in a simulated environment, at Kirkpatrick evaluation level two (change in skills), with only a few situated in authentic clinical practice at Kirkpatrick level three (change in skills applied at work), which may limit application to routine clinical practice.[72]

Implications for future research and clinical practice

Our review supports the likely beneficial impact of verbal face-to-face feedback on health professionals' task performance, compared with no feedback. By analysing included studies based on factors known to influence performance, our review assists future researchers by clarifying key parameters that need to be considered. Many of the included studies were 'one-off', involved small numbers of participants and included sources of bias. This indicates the need for studies that involve more participants and are methodologically better designed and executed. In addition, to address publication bias, larger published studies or identification of unpublished studies are needed. To advance this field of knowledge, research programmes designed to systematically investigate the constituents required for effective feedback are needed. This is likely to involve a series of studies designed to isolate one factor at a time, with all other key influences on performance standardised, in order to identify and replicate the conditions that are most effective in helping learners to improve, across different contexts. As key elements in effective feedback are established, implementing this knowledge across health professions education will be important to optimise both clinical practice and patient outcomes.

SUMMARY

We systematically collated the available evidence regarding the impact of verbal face-to-face feedback on health professionals' workplace task performance, compared with no or alternative feedback. In a meta-analysis, we found that verbal face-to-face feedback may result in a moderate to large improvement in workplace task performance, compared with no feedback SMD 0.7 (95% CI 0.37 to 1.03; p<0.001), after excluding studies at high risk of bias. We extracted and reported data on factors known to influence performance development, which included both components within feedback interventions and additional factors, such as providing teaching or practice opportunities. The diversity in feedback interventions identified in this review (even within 'face-to-face feedback') highlights the need to view feedback as a complex intervention.

Author affiliations
[1]Monash Doctors Education, Monash Health; Faculty of Medicine, Nursing and Health Sciences, Monash University; Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia

[2]Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia
[3]Monash Health, Melbourne, Victoria, Australia
[4]Department of Physiotherapy, Monash University, Clayton, Victoria, Australia

Twitter Christina Elizabeth Johnson @chrisj_au

Acknowledgements The authors wish to thank Anne Young, Subject Librarian, Hargrave-Andrew Library, Monash University for her assistance with the electronic databases searches.

Contributors CEJ and JLK designed the review; CEJ created the search strategy and undertook the literature searches; CEJ and MPW undertook study screening, data extraction and quality assessment with oversight from JLK; CEJ and JLK undertook data analysis and interpreted the findings; CEJ prepared the manuscript; and all authors contributed to the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data access can be requested by contacting the corresponding author.

ORCID iDs
Christina Elizabeth Johnson http://orcid.org/0000-0002-4209-8419
Mihiri P Weerasuria http://orcid.org/0000-0001-6897-3982
Jennifer L Keating http://orcid.org/0000-0003-3161-4964

REFERENCES

1 Morris C, Blaney D. Work-based learning. In: Swanwick T, ed. *Understanding medical education: evidence, theory and practice*. 2nd ed.. Oxford: The Association for the study of Medical Education, 2014: 97–109.
2 Carraccio C, Englander R, Van Melle E, *et al*. Advancing competency-based medical education: a charter for clinician-educators. *Acad Med* 2016;91:645–9.
3 van der Vleuten CPM, Schuwirth LWT, Driessen EW, *et al*. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34:205–14.
4 Holmboe ES. Realizing the promise of competency-based medical education. *Acad Med* 2015;90:411–3.
5 Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA* 2009;302:1316–26.
6 Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;44:101–8.
7 Bing-You R, Hayes V, Varaklis K, *et al*. Feedback for learners in medical education: what is known? A scoping review. *Acad Med* 2017;92:1346–54.
8 Johnson CE, Keating JL, Boud DJ, *et al*. Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. *BMC Med Educ* 2016;16:96.
9 Watling CJ, Ginsburg S, Assessment GS. Assessment, feedback and the alchemy of learning. *Med Educ* 2019;53:76–85.
10 Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol* 2002;57:705–17.
11 Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad Med* 2015;90:609–14.
12 Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Med Teach* 2011;33:961–73.
13 Johnson CE, Keating JL, Farlie MK, *et al*. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. *BMC Med Educ* 2019;19:129.
14 Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;77:81–112.
15 Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008;15:988-94.
16 Molloy E, Boud D. Changing conceptions of feedback. In: Molloy E DB, ed. *Feedback in higher and professional education*. London: Routledge, 2013: 11–33.
17 Ende J. Feedback in clinical medical education. *JAMA* 1983;250:777–81.
18 Veloski J, Boex JR, Grasberger MJ, *et al*. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME guide No. 7. *Med Teach* 2006;28:117–28.
19 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012:CD000259.
20 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–-9.
21 Higgins JPT, Churchill R, Chandler J, *et al*. Cochrane Handbook for systematic reviews of interventions version 5.2.0 (updated June 2017), 2017. Available: www.training.cochrane.org/handbook
22 Sterne JAC, Savović J, Page MJ, *et al*. Rob 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
23 Verhagen AP, de Vet HCW, de Bie RA, *et al*. The Delphi list. *J Clin Epidemiol* 1998;51:1235–41.
24 Sterne JAC, Sutton AJ, Ioannidis JPA, *et al*. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.
25 Sterne JAC, Egger M, Moher D, *et al*. Chapter 10: Addressing reporting biases. In: Higgins JPT, Churchill R, Chandler J, *et al*, eds. *Cochrane Handbook for systematic reviews of interventions version 520 (updated June 2017*, 2017. www.training.cochrane.org/handbook
26 Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549.
27 Higgins JPT, Green S. Cochrane Handbook for systematic reviews of interventions version 5.1.0, 2011. Available: www.handbook.cochrane.org [Accessed Mar 2011].
28 Schünemann HJ, Oxman AD, Vist GE, *et al*. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Churchill R, Chandler J, *et al*, eds. *Cochrane Handbook for systematic reviews of interventions version 520 (updated June 2017*, 2017. www.training.cochrane.org/handbook
29 Grade guidelines Journal of clinical epidemiology series. Available: https://www.gradeworkinggroup.org/2020
30 Ahlborg L, Weurlander M, Hedman L, *et al*. Individualized feedback during simulated laparoscopic training:a mixed methods study. *Int J Med Educ* 2015;6:93–100.
31 Boyle E, O'Keeffe DA, Naughton PA, *et al*. The importance of expert feedback during endovascular simulator training. *J Vasc Surg* 2011;54:240–8.
32 Olms C, Jakstat HA, Haak R. The implementation of Elaborative feedback for qualitative improvement of shade Matching-A randomized study. *J Esthet Restor Dent* 2016;28:277–86.
33 Pavo N, Goliasch G, Nierscher FJ, *et al*. Short structured feedback training is equivalent to a mechanical feedback device in two-rescuer BLS: a randomised simulation study. *Scand J Trauma Resusc Emerg Med* 2016;24:70.
34 Bonrath EM, Dedy NJ, Gordon LE, *et al*. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. *Ann Surg* 2015;262:205–12.
35 Kroft J, Ordon M, Po L, *et al*. Preoperative practice paired with Instructor feedback may not improve obstetrics-gynecology residents' operative performance. *J Grad Med Educ* 2017;9:190–4.
36 Soucisse ML, Boulva K, Sideris L, *et al*. Video coaching as an efficient teaching method for surgical Residents-A randomized controlled trial. *J Surg Educ* 2017;74:365–71.
37 Xeroulis GJ, Park J, Moulton C-A, *et al*. 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* 2007;141:442–9.
38 O'Connor A, Schwaitzberg SD, Cao CGL. How much feedback is necessary for learning to suture? *Surg Endosc* 2008;22:1614–9.
39 Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *Am J Med* 1983;75:465–70.

40 Vafaei A, Heidari K, Hosseini M-A, *et al*. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emerg* 2017;5:e28.
41 Guyatt G, Oxman AD, Sultan S, *et al*. Grade guidelines: 11. making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151–7.
41 Al-Jundi W, Elsharif M, Anderson M, *et al*. A Randomized Controlled Trial to Compare e-Feedback Versus "Standard" Face-to-Face Verbal Feedback to Improve the Acquisition of Procedural Skill. *J Surg Educ* 2017;74:390–7.
41 Baldwin LJL, Jones CM, Hulme J, *et al*. Use of the learning conversation improves instructor confidence in life support training: an open randomised controlled cross-over trial comparing teaching feedback mechanisms. *Resuscitation* 2015;96:199–207.
41 Boehler ML, Rogers DA, Schwind CJ, *et al*. An investigation of medical student reactions to feedback: a randomised controlled trial. *Med Educ* 2006;40:746–9.
41 Bosse HM, Mohr J, Buss B, *et al*. The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Med Educ* 2015;15:22.
41 Brinkman WB, Geraghty SR, Lanphear BP, *et al*. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med* 2007;161:44–9.
47 DeLucenay AJ, Conn KM, Corigliano A. An evaluation of the impact of immediate compared to delayed feedback on the development of counselling skills in pharmacy students. *Pharmacy Education* 2017;17:322–8.
47 Lee JY, McDougall EM, Lineberry M, *et al*. Optimizing the timing of expert feedback during simulation-based spaced practice of endourologic skills. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 2016;11:257–63.
47 Ozcakar N, Mevsim V, Guldal D, *et al*. Is the use of videotape recording superior to verbal feedback alone in the teaching of clinical skills? *BMC Public Health* 2009;9:474.
47 Strandbygaard J, Bjerrum F, Maagaard M, *et al*. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. *Ann Surg* 2013;257:839–44.
47 van de Ridder JMM, Peters CMM, Stokking KM, *et al*. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Adv Health Sci Educ Theory Pract* 2015;20:803–16.
52 van de Ridder JMM, Berk FCJ, Stokking KM, *et al*. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Med Teach* 2015;37:767–74.
52 Rogers DA, Boehler ML, Schwind CJ, *et al*. Engaging medical students in the feedback process. *Am J Surg* 2012;203:21–5.
52 Manzone J, Tremblay L, You-Ten KE, *et al*. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Med Educ* 2014;48:430–40.
52 Sox CM, Dell M, Phillipi CA, *et al*. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134:965–71.
56 Backstein D, Agnidis Z, Sadhu R, *et al*. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Can J Surg* 2005;48:195–200.
56 Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum Associates, 1988.
58 Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;119:254–84.
59 Shute VJ. Focus on formative feedback. *Rev Educ Res* 2008;78:153–89.
58 Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Med Educ* 2016;50:933–42.
61 Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Med Teach* 1997;19:40–4.
62 Sargeant J, Mann K, Sinclair D, *et al*. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ* 2007;41:583–91.
61 Eva KW, Armson H, Holmboe E, *et al*. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and Reasoning processes. *Adv Health Sci Educ Theory Pract* 2012;17:15–26.
64 Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Med J Aust* 2004;181:327–8.
65 Fitts PM, Posner MI, Performance H. *Belmont Ca*. Brooks/Cole, 1967.
66 Li Q, Ma E-L, Liu J, *et al*. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Med Teach* 2011;33:e549–55.
66 Elliott SN. *Educational psychology: effective teaching, effective learning*. 3rd edn. Boston: McGraw-Hill, 2000.
66 Wadsworth BJ. *Piaget's theory of cognitive and affective development: Foundations of constructivism*. 5th edn. White Plains: Longman Publishing, 1996.
69 Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. *Educ Psychol Rev* 1998;10:251–96.
69 Kaufman DM, Mann KV. Teaching and learning in medical education: How theory can inform practice. In: Swanwick T, ed. *Understanding medical education evidence, theory and practice*. 2nd edn. Oxford: Wiley Blackwell, 2014: 7–29.
71 Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychol Inq* 2000;11:227–68.
72 Belfield C, Thomas H, Bullock A, *et al*. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;23:164–70.
72 Rudolph JW, Simon R, Dufresne RL, *et al*. There's No Such Thing as "Nonjudgmental" Debriefing: A Theory and Method for Debriefing with Good Judgment. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 2006;1:49–55.