BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030672
Article Type:	Research
Date Submitted by the Author:	28-Mar-2019
Complete List of Authors:	Johnson, Christina; Monash Health, Monash Doctors Education ; University of Melbourne, Department of Medical Education, Melbourne Medical School Weerasuria, Mihiri; Monash Health Keating, Jennifer; Monash University, Department of Physiotherapy
Keywords:	MEDICAL EDUCATION & TRAINING, feedback, health professions education, formative feedback

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Short title: Systematic review and meta-analysis on feedback

Christina E. Johnson MBChB, MHPE, FRACP

ORCID ID: 0000-0002-4209-8419

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, Australia.

Mihiri P. Weerasuria MBBS (Hons)

ORCID ID: 0000-0001-6897-3982.

Monash Health, Melbourne, Australia.

Jennifer L. Keating PhD, GradDipManip PT, BAppSci (PT)

ORCID ID: 0000-0003-3161-4964

Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia.

Correspondence should be addressed to Christina Johnson, Monash Doctors Education, McCulloch House, Monash Health, Clayton Road, Clayton, Victoria 3168, Australia. Telephone: 613-9594 3743. Email: 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 created a search strategy using MEDLINE subject headings and key words related to 'feedback' AND 'health professional' AND 'performance' AND 'randomised controlled trial'. The search strategy was adapted for other databases. Ovid MEDLINE, CENTRAL, Embase, CINAHL and PsychINFO were searched until 1st February 2019. Two authors independently undertook study selection. 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 to no feedback or alternative feedback, and available as full text publications in English. One author extracted data using a pre-piloted standardised form and another author checked the accuracy. The quality of evidence was assessed using the GRADE criteria. For feedback compared to no feedback, outcome data from included studies was pooled using a random effects model.

Results: For verbal face-to-face feedback compared to no feedback, eight higher quality studies involving 392 health professionals were included in a meta-analysis: the standardised mean difference (SMD) was 0.7 (95% CI 0.37-1.03; P<0.0001) in favour of feedback. The calculated SMD prediction interval was -0.06 to 1.46. Potentially important variation in

feedback source, timing and content were identified. For feedback compared to alternative feedback, studies could not be pooled due to substantial intervention and design heterogeneity.

Conclusions: Verbal face-to-face feedback in the health professions substantially enhances workplace task performance, compared to no feedback.

PROSPERO 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 to no feedback is the first to provide an estimate of the benefit of verbal feedback on performance of a workplace task in the health professions.

• For the meta-analysis, not all studies, that met the inclusion criteria, clearly reported the data required for pooling, so for some studies it was obtained from the author or estimated by calculating it using available data or reading off graphs; otherwise, the study was excluded.

Keywords: feedback, effective feedback, formative feedback, systematic review, meta-analysis, health professions education

INTRODUCTION

Health professions education is embedded in clinical practice for both students and qualified staff as they continue learning and training.[1] A common form of feedback in this context is 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). This may occur informally 'on the run' during routine clinical practice[2] or in a more formalised way, for example as a workplace-based assessment, or at the end of a day or the end of a clinical placement.[3-5]

There is widespread acceptance that feedback has an important role in maximising learning and achievement.[6-8] The value of feedback for a learner lies in the opportunity to enhance their understanding of how the task should be done and how they can improve their performance.[9] Feedback can occur in various forms, including verbal, written or automated (for example, from a simulator or within an online learning module). The unique potential benefit of face-to-face verbal feedback is the opportunity for i) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to co-construct insights and strategies for improvement and ii) inter-personal 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.[10-13]

Within the health professions, few previous systematic reviews have focused on feedback. In 2006, Veloski et al[14] 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 commonest data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers et al[15] updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared to usual care (median risk difference (RD) 4.3% (IQR 0.5-16%). The review included various qualified health professionals but 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 was small at 4.3%, the 75th centile was 16%, which suggested that audit and feedback interventions could be much more effective. Using multivariable meta-regression, they identified 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 to just information on performance (RD 5%). In addition, two other factors increased the likelihood of improvement: a lower baseline performance and the type of behaviour being targeted e.g. prescribing (possibly perceived as 'important' and 'straightforward') had better outcomes than improving diabetes management (more complex) or test ordering (possibly perceived as 'less important').

No systematic review has investigated the impact of verbal face-to-face feedback on a health professional's performance, the typical scenario in clinical practice environments. Therefore, our research question for this systematic review was

'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 the interventions and outcomes reported by included studies.

METHODS

Inclusion and exclusion criteria

This review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) Registration ID CRD42017081796. Studies were eligible for inclusion if they were randomised controlled trials in which individual health professionals were randomised to feedback, compared to no feedback or alternative feedback. A broad definition of feedback was permitted with a minimum requirement that it included information regarding learner performance. At least one intervention had to involve verbal face-to-face feedback from a health professional, based on the observable performance of a workplace task performed by another health professional. Performance following interventions had to be objectively assessed. To isolate the effects of feedback, other conditions had to be comparable for both groups. Participants had to be 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. Reports had to be available as full text publications in English. Studies were excluded if a health professional delivered pre-determined comments or if feedback was solely provided by a simulated patient or machine. Audit and feedback studies, where feedback was based on aggregated quality performance data (such as numbers of tests ordered or compliance with standards) were excluded. Studies were also 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.

Literature search strategy and article screening

We developed the search strategy in collaboration with a senior medical librarian using MEDLINE subject headings and key words, including synonyms, truncation, wildcard and proximity operators related to 'feedback' AND 'health professional' AND 'performance' AND 'randomised controlled trial' were used (see Appendix 1 for the full search strategy for OvidMEDLINE). 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 and PsychINFO were searched until 1st February 2019. We also searched the reference lists of systematic reviews and included studies to identify relevant studies.

One review author (CJ or MW) screened titles to exclude clearly irrelevant reports. Two authors (CJ and MW) independently screened remaining abstracts to identify potential eligible studies, then independently assessed the full text of these studies. Decisions were compared using Covidence, an on-line platform, and disagreements were resolved through discussion, including a third review author (JK) if necessary.

Data extraction

One review author (CJ) used a pre-piloted standardised form to extract data from included studies and another author (MW or JK) independently checked it. We resolved discrepancies through discussion.

GRADE quality of evidence assessment

Two authors (CJ and JK) independently assessed quality of evidence for each study outcome using the GRADE approach[16] based on information extracted by one author (CJ) and checked

by another (MW). Concerning the risk of bias assessment, in order to clarify our interpretation of the general advice included in the Cochrane GRADE document,[17] we used the following decision rules. Throughout 'unclear risk' was ascribed the same risk as 'high risk', as this was deemed to indicate a substantial risk of bias. For 'random sequence generation': studies were rated as 'low risk' when it was stated that participants were enrolled then 'randomised' or specified a valid randomisation method. For 'allocation concealment': a 'low risk' rating was only given if studies explicitly stated the method used to conceal which group the potential participants would be allocated to, to prevent anyone influencing this. If this information was missing or inadequate, the study was rated as 'unclear'. For 'participant and research team blinding': both participants and researchers were required to be effectively blinded to be rated as 'low risk'. A participant receiving feedback or a researcher giving feedback was deemed not to be blinded, even if they were deliberately not informed about the intervention or any differences between interventions. If the information was insufficient or ambiguous, this was rated as 'unclear risk'. If available information indicated that either the participant or research team members were not blinded, this was rated as 'high risk'. For 'outcome assessor blinding': if there no explicit statement that outcome assessors were blinded, it was rated as 'unclear'. If there was information that implied an outcome rater had an opportunity to know which group a participant was allocated to, this was rated as 'high risk'. 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[18]), and to clearly state the participant numbers at the start and the number that dropped out during the study, from which group and the reasons. If this information could be derived from the presented information, this was similarly acceptable. For 'selective outcome reporting': to be rated as 'low risk', studies were required to clearly present outcome data (including for each group, the number of participants, mean and standard deviation, or similar) in accordance with a prior published

protocol or, if there was no prior published protocol, to explicitly state there had been no change in the research protocol after commencement of the study. Studies that did not have a pre-published protocol but clearly reported data as expected and outlined in the methods were rated as 'unclear'.

We defined 'higher quality studies' for meta-analysis as those studies that had no 'high' risk ratings and a maximum of three 'unclear' ratings that did not involve 'sequence generation', 'outcome assessor blinding' or 'incomplete outcome data'. This played down the potential risks of bias associated i) with 'participant and research team blinding' which was not possible due to the nature of the feedback intervention, and ii) with 'selective outcome reporting' as many studies did not have pre-published protocols but did report outcomes as expected and in accordance with the outlined methods.

Data analysis

Outcomes from included studies were expressed using point estimates and measures of variability (for example means (standard deviations SD) or medians (interquartile range IQR)). The effect was quantified using the standardised mean difference to combine studies measuring the same outcome (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 pooled outcomes for comparable comparisons in a meta-analysis using RevMan software (Review Manager Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). We used a fixed-effect model and assessed heterogeneity. As a fixed effect model assumes that all studies are measuring the same 'treatment effect' with the variability explained by chance,[19] where tests for heterogeneity

returned a p value <0.1 or I^2 > 25%, we repeated the meta-analysis using a random-effects model.[19,20] Finally, we repeated the meta-analysis using only higher quality studies and calculated a prediction interval. The standardised mean difference and 95% CI is the average effect across multiple studies and its error estimates,[20] whereas a prediction interval describes the range of likely results for new individual studies.

Patient and Public Involvement

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

RESULTS

Literature search

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.

[Figure 1 here]

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

Included studies

Eleven randomised controlled trials investigated the effect of verbal face-to-face feedback compared to no feedback on the objective evaluation of a workplace task (see Table 1 which describes the studies included in the comparison of feedback to no feedback). Seven (64%) reports were published in the five years since 2014. The studies were conducted in Europe (4),[21-24] Canada (4),[25-28] the USA (2)[29,30] and Asia (1).[31]

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

Author Year Country	Task	Participants Health Profession % Male	Teaching and Practice Same for Feedback Intervention and Control groups	Feedback Intervention Additional information	Source		Timing			Content				
					Subject Expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info[a]	Performance video	Written performance info
Ahlborg 2015 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 x practice trials. Performance evaluation: end of session.	2 x fb episodes. Fb given by expert i) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus ii) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓		
Bonrath 2015 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.) x 25 min fb episodes. Fb given by expert using specific coaching model[b] + video review of learner operating + video exemplars of good/poor technique. Effectiveness of strategies reviewed at subsequent session.	✓				✓	✓	✓		✓	
Boyle 2011 (expert fb)	Simulated endovascular surgery using a VR simulator	Doctors training in surgery PGY4+	Intervention duration: 1 session Teaching + expert demonstration. 5 x practice trials	5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' + simulator output information.	✓			✓		?	?	✓		

Ireland	(renal artery angioplasty + stent)		Performance evaluation: end of session										
Boyle 2011 (peer fb)	Same as above	Same as above	Same as above	5 x fb episodes. Peer discussed simulator output, any task errors & teaching instructions given at start.		✓		✓		✓	✓	✓		
Kroft 2017 Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy)	Doctors training in O&G PGY2-6 33% M	Intervention duration: 1 x 15min practice using laparoscopic salpingectomy module on VR surgical simulator within 1h of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1 x fb episode from expert directly after VR simulator practice. Fb 'standardized and given in an evidence based fashion to optimise effectiveness' & included '3 constructive recommendations based on performance'.	✓			✓		✓	✓			
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	Medical students UGY1-2 44% M	Intervention duration: 4 wk. 2h instruction + practice suturing & knot tying until able to do it easily. Then instruction on laparoscopic surgery + expert demonstration video of task tying, followed by 30 mins familiarisation with equipment. Practice: 1h 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.	✓		✓	✓		✓	✓	✓		
Olms 2016 Germany	Simulated colour matching teeth	Dental students UGY3	Intervention duration: 1 session Study conducted during 10 wk routine university module on matching tooth shades involving variety of teaching + practice opportunities.	1 x expert fb session. Fb included correct response + explanation with expert demonstration if needed + written copy of evaluation. Expert trained in fb.	✓			✓		✓	✓			✓

			Performance evaluation: 2 wks after intervention (within one university module).										
Pavo 2016 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 x 2h additional training session: instructional video + training using modified Peyton 4 step approach.[c] Brief practice (few mins) 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 & depth, hand position, decompression & hands-off time. Instructional video for intervention group had demonstrated this.		✓	✓			✓	✓			
Skeff 1983 USA	Clinical teaching skills during ward round	Physicians	Intervention duration:1 session in the middle of 4wk ward duty. At mid & 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 wk later, at end of ward duty	1 x 60 min fb discussion with peer, including video review, trainee ratings & self-assessment to enable physician to identify strengths & devise solutions to problems.	✓				✓	✓	✓		✓	✓
Soucisse 2017 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 wk later (ongoing clinical work as a surgical resident).	1 x 30min expert fb sometime after baseline performance with video review of baseline performance + coaching using 'GROW' model[d] including 2-3 suggestions for improvement + expert demonstration followed by learner demonstration of desired improvements, as required + action plan.	✓				✓	✓	✓		✓	

Vafei 2017 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 x 5min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓				
Xeroulis 2007 (fb after) Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	Medical students UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. Performance evaluation: end of session.	Expert fb as needed (expert or learner initiated), after practice trials, involving constructive ways to improve + expert demonstration.	✓			✓		✓	✓			
Xeroulis 2007 (fb during)	Same as above	Same as above	Same as above	Same as above except expert fb during practice trials.	✓		✓			✓	✓			

Participants and workplace tasks

Included studies involved 488 participants of which 290 (60%) were medical students in four studies,[21,24,28,29] 138 (28%) were doctors (doctors training in surgery in three studies,[22,25,27] training in obstetrics and gynaecology in one study[26] and training in emergency medicine in one study,[31] and physicians in one study[30]) and 60 (12%) were dental students in one study.[23] In all studies, the performance evaluated was a discrete task; there were no longitudinal evaluations. The task occurred in a simulation setting in seven studies (7/11, 64%) and in clinical practice in four studies (4/11, 36%). The task was a surgical procedure in seven studies (7/11, 64%). Five studies involved simulated surgical tasks including bench top models for knot tying[28] and forming a bowel anastomosis;[27] using a laparoscopic simulator for suturing and knot tying;[29] and using a virtual reality (VR) simulator for laparoscopic surgery[21] and endovascular surgery.[22] Two studies involved laparoscopic surgery in clinical practice.[25,26] The remaining four studies evaluated simulated matching of tooth colour in a dental school,[23] simulated cardiopulmonary resuscitation (CPR),[24] chest ultrasound for emergency trauma patients[31] and teaching skills in clinical practice.[30]

Heterogeneity amongst verbal face-to-face feedback interventions

Although all studies satisfied inclusion criteria, there were variations in feedback components that have been reported to influence outcomes (see Table 1).[14,15] These differences included feedback source (expert or peer), timing (during, directly after the performance or delayed) and content (evaluative performance information from a health professional or with the addition of written information, output information from a simulator or CPR machine, or advice on how to improve performance).

There were also diverse differences in additional factors previously reported to modify performance including participants' prior experience, provision of instruction and expert demonstration of the task, and practice.[14,15] The participants were novices to the assessed task in five studies (5/11, 45%);[21-23,28,29] the participants had prior experience in six studies.[24-27,30,31] The intervention included instruction and expert demonstration of the task in six studies (6/11, 55%): these included all five studies involving novice participants[21-23,28,29] plus one study that involved CPR for medical students, many of whom had attended a course previously.[24] The other five studies involved doctors working in clinical practice; in these studies, no instruction or expert demonstration was included within the research intervention but may or may not have occurred during the routine course of their work during that time. One study involved physicians' teaching on ward rounds[30] and the other four studies assessed tasks by doctors training in relevant specialties.[25-27,31] The amount of practice varied substantially between different studies, for both simple and complex tasks. For example, comparing two studies that involved simple surgical knot tying: in Xeroulis,[28] participants practiced 18 times in one session and in O'Connor, [29] they could practice up to an hour a day, for 24 days. Looking at more complex surgical procedures, such as simulated surgery using a virtual reality (VR) simulator: in Ahlborg,[21] participants had two practice attempts at the simulated surgery (laparoscopic salpingectomy) and in Boyle,[22] participants had five attempts at the simulated surgery (renal artery angioplasty and stenting) before the performance evaluation.

The intervention period ranged from one day (most common) up to two months.[25] Nine (9/11, 82%) studies involved a single session (involving one episode of feedback in five studies[23,26,27,30,31] and multiple episodes of feedback in four studies[21,22,24,28]). Two studies (2/11, 18%) involved multiple feedback sessions over a longer period: one study[25] included approximately four coaching sessions regarding bariatric surgery across a two month surgical attachment, and another[29] included almost daily one hour practice sessions for laparoscopic

suturing, with feedback throughout each one, over four weeks. Also, the timing of the post-feedback performance assessment, in relation to the intervention, differed. It occurred directly following the intervention in seven studies: at the end of the single session in five studies[21,22,24,26,28] and at the end of an extended intervention period in two studies.[25,29] In the other four studies, the post-feedback performance assessment occurred some weeks after the intervention was completed but while relevant exposure to possible teaching and/or practice opportunities continued. Olms[23] included a single feedback session, with the final evaluation two weeks later, in the midst of a routine one month university teaching unit on tooth shade matching. Skeff[30] arranged a single coaching session on ward round teaching in the middle of physicians' four week ward duty, with the final evaluation post-performance evaluation at the end. Soucisse[27] also organised a single coaching session for surgical residents, with the final evaluation occurring three weeks later. Vafaei[31] involved a single workplace-based assessment with feedback for doctors training in emergency medicine on chest ultrasound for emergency trauma patients, followed by a two month period of routine clinical work before the post-feedback assessment.

Baseline performance

Although randomisation removes the need to check comparability in baseline task performance for intervention and comparison groups, it may be useful to check this when participant numbers are small. In addition, previous research reported that improvement was more likely when baseline performance was low.[15] Seven studies reported no statistically significant differences between baseline performances for the comparison groups.[23,27,28,30,31] and four studies did not report baseline task performance.[21,22,24,29]

Research funding

Regarding research funding, one study[24] that focused on cardiopulmonary resuscitation (CPR) quality, was loaned a device (used to measure CPR parameters and provide automated feedback

to participants) for the period of the study by Philips but the company was not otherwise involved in the research; five studies received funding from independent institutions,[21,25,28,30] three studies did not receive any funding[22,23,31] and two studies did not report information on funding.[27,29]

Meta-analysis of verbal face-to-face feedback compared to no feedback

This meta-analysis included 13 comparisons from the 11 studies, involving 488 participants. Two studies reported data that each enabled two comparisons: in one study, feedback provided *during* practice in one group and *directly after* practice in another was compared to no feedback;[28] in another study, feedback provided by an *expert* in one group and by a *peer* in another group[22] was compared to no feedback. In the meta-analysis, numbers for the control group for these studies were halved to retain sample independence.[32]

Meta-analysis of the effect of verbal face-to-face feedback compared to no feedback on workplace task performance found a standardised mean difference of 1.09 (95% confidence interval (CI) 0.59-1.59; P=0.00001), using a random effects model (as I^2 =78%) (see Figure 2).

As a sensitivity analysis, we repeated the meta-analysis with only the higher quality studies. Based on risk of bias assessment, eight studies (8/11, 73%) were included.[21,22,24-28,30] This random effects analysis (I^2=34%), which involved 392 health professional learners across ten comparisons, resulted in a standardised mean difference of 0.7 (95% CI 0.37-1.03; P<0.0001) and prediction interval of -0.06 to 1.46. One study[24] had a dominant effect on results because it contributed over half the pooled participant numbers in the meta-analysis (224/392, 57%). This can be clearly seen on the funnel plot for the higher quality studies (see Figure 3b), in which this study is represented by the highest circle on the left-hand side of the vertical midline.

Comparison 1: GRADE quality of evidence assessment

Risk of bias assessment

The risk of bias ratings for each study are shown in Figure 2. All included studies were confirmed to be appropriately randomised. Only four studies (4/11, 36%) clearly explained an effective strategy to conceal group allocation of participants.[24-27] The participants and research team members were not blinded in any included studies because the intervention involved feedback between a research team member and a participant, consistent with most education interventions. The post-intervention performance assessment was assessed as low risk in eight studies because it was from either blinded assessors who rated videos of the participants' performance[22,25-28,30] or by a machine (simulator or CPR machine).[21,24] However, in three studies, the feedback provider and outcome assessor appeared to be the same person, so these were rated as 'high risk'.[23,29,31] Across all the studies, the follow up rate for each group was at least 85%. Only two studies had a prior published protocol in addition to reporting all outcomes as planned.[25,27] For all other studies, it could not be ascertained if outcomes had been selectively reported but outcomes were reported (or the outcomes could be derived by reading it off from a graph or deriving it from other data) as expected and as outlined in the methods, so these were rated as 'unclear', except one. This one study was rated as 'high risk' for selective outcome reporting because it did not include the expected information on performance post-intervention.[29]

GRADE quality assessment

We judged that the outcome of 'objective assessment of a health professional's performance' was at high risk of bias across multiple criteria in included studies and the overall body of evidence indicated this was likely to seriously alter the results, so we downgraded the overall evidence by one level. The two aspects that were most influential on our decision were the lack of allocation concealment and prior published protocols. Participant and research team member blinding was not possible due to the intervention. However, this had limited impact on the

selected outcome 'objective assessment of performance', as long as the outcome assessment was made by blinded assessors or machines, which was the case for all but three studies (8/11, 73%).[23,29,31] Regarding the other criteria in the GRADE assessment of quality, we judged there to be some variability across study results and therefore downgraded the evidence due to inconsistency by one level. This was based on the methodological and statistical heterogeneity which was not explained by subgroup analysis. We judged the results to be directly applicable to our review question and therefore the evidence was not downgraded for indirectness. We judged the effect size to be sufficiently precise and therefore did not downgrade the evidence for imprecision. This was based on sufficient numbers of participants (488 participants with all studies included and 392 with just higher quality studies included[21,22,24-26,28,30,31] and a consistent beneficial effect, indicated by the confidence interval for the overall effect estimate not crossing zero and all individual studies showing a beneficial effect with substantial overlap in their confidence intervals. Finally, we judged that there was likely to be a systematic overestimation of the underlying beneficial effect of feedback because we strongly suspected publication bias and therefore we downgraded the evidence by one grade. The funnel plots (see Figure 3), particularly the one including only higher quality studies (see Figure 3b) is asymmetrical, with a paucity of small studies with negative effect sizes that are less likely to be published. However, funnel plots may not be appropriate for assessing publication bias when there are a small number of studies[33] especially when there is high heterogeneity, as the interpretation of a funnel plot is based on the assumption that the intervention effect is the same across studies.[32]

[Figure 3 here]

Summary of GRADE quality assessment

In summary, combining all five GRADE criteria for assessing the quality of evidence, we downgraded the overall rating by one, from high to moderate. We judged that the quality of the

evidence contributing to the effect estimate of 0.70 in the comparison of verbal face-to-face feedback to no feedback involving only the eight higher quality studies, was moderate. Hence, we are moderately confident in the effect estimate; the true effect is likely to be close to this but there is a possibility that it is substantially different. Figure 4 displays the Summary of Findings table for the meta-analysis of the effect of feedback, compared to no feedback, on the workplace task performance of a health professional.

[Figure 4 here]

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

Participants and assessment tasks

All 20 studies (22 comparisons) in this analysis included at least one intervention involving verbal, face-to-face feedback compared to alternative feedback (see Table 2 which describes the studies included in the comparison of verbal face-to-face feedback compared to alternative feedback). Nine studies (9/20, 45%) were published in 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%). Across these studies there were 1974 participants, including medical students (1076, 55%) in 14 studies,[24,28,29,34-44] mixed health professional students (640, 32%) in one study,[45] doctors (105, 5%) in four studies,[22,30,46,47] and pharmacy students (153, 8%) in one study.[48] All studies included assessment of a discrete task except for two studies which involved longitudinal evaluations.[30,47] Three studies evaluated performance in a clinical practice setting (involving teaching skills[30] professional and communication skills[47] and oral case presentations[42]) and the remaining 17 assessed performance of tasks in a simulated environment.[22,24,28,29,34-41,43-46,48] Simulated surgical tasks included suturing and/or knot tying,[28,29,34-36] bench top surgical procedures such as vascular anastomosis,[46] flexible

ureteroscopy for urolithiasis,[38] renal artery angioplasty and stent,[22] or surgery using a VR simulator for a laparoscopic salpingectomy.[44] Simulated critical care tasks included basic life support(BLS)/CPR[24,45] and intubation.[39] The remaining simulated tasks included a hearing test,[40,41] simulated patient consultation[43] and nasogastric tube insertion.[37]

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

Article First author Year Country	**Task**	**Participants:** Health profession Gender: % Men	**Common to interventions A + B**	**Intervention A** All verbal face-to-face feedback to an individual health professional	**Intervention B**
Al-Jundi[a] 2017 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 x 10 mins for baseline performance. Performance evaluation: 2 days later	**Immediate face-to-face + written expert feedback** 1 x 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 x 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.
Backstein 2005 Canada	Simulated surgical procedure using a bench top model (vascular anastomosis)	**Doctors in surgical training** PGY1	Intervention duration:4 wk Lecture on surgical procedure. 3 x 2h 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 wk 4	**Review of performance video with expert fb** **+ practice sessions with expert fb available** 3 x weekly videotaping of surgical procedure, with expert feedback available during task, followed by up to 15min review of video with expert fb	**Practice sessions with expert fb available**
Baldwin 2015 England	Simulated BLS	**Health profession students** medical (58%), physiotherapy (12%), pharmacy (10%), nursing	Intervention duration: 4 wk Instruction and practice with manikin 3 x 2.5h weekly. Fb provided directly after performance by senior peer instructor. Instructor accredited in	**'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[b], with final summary.	**'Feedback sandwich model'** Fb involved a point for improvement in between 2 points of praise.

		(10%), dentistry (10%) UGY1 33% M	BLS + trained to provide fb. Fb provider compliance monitored. Performance evaluation: in wk 4		
Boehler 2006 USA	Simulated surgical skill using a bench top model (tying a 2-handed square knot)	**Medical students** UGY2-3 52% M	Intervention duration: 1 session Instruction in knot tying from surgeon. 1 x baseline performance. Performance evaluation: end of session.	**Expert feedback** 1 x episode of fb from expert surgeon, directly after performance, describing 1-2 specific ways to improve performance.	**Compliment** 1 x pre-scripted general compliment e.g. 'great job!'
Bosse 2015 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 4 step procedural training method[c] 6 x practice trials. Fb 'positively worded', focused on effect of actions, given directly after performance by senior peer instructors, trained in procedure & fb. Performance evaluation: end of session.	**High frequency fb** 6 x episodes of fb, given after each practice trial.	**Low frequency practice** 2 x episodes of fb, given after first and last practice trial.
Boyle 2011 Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty + stent)	**Doctors training in surgery** PGY4+	Intervention duration: 1 session Teaching & expert demonstration. Fb providers had simulator training. 5 x practice trials (each maximum 40min). Performance evaluation: end of session.	**Expert fb** 5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	**Peer fb** 5 x fb episodes. Peer discussed simulator output, any task errors & the teaching instructions given at start of session.
Brinkman 2007 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 feedback 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.	**Coaching session + routine feedback as part of clinical training** 1 x 30min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses & parents. Used a coaching model[d] focused on assisting learner to understand information, design goals and improvement strategies. Fb given	**Routine feedback as part of clinical training** Performance ratings from nurses and patients not seen.

			Performance evaluation: 5 months after start of clinical attachment.	by paediatricians trained in coaching model.	
DeLucenay 2017 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 5min fb on communication skills. Performance evaluation: last 4 SP consultations.	**Immediate face-to-face fb** 4 x 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 x videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.
Lee 2016 Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis)	**Medical students** UGY3-4 78% M	Intervention duration: 3 wk Instruction and expert demonstration of procedure, followed by 3 x weekly 30min practice sessions. Performance evaluation: end of 3rd session.	**Early feedback** 1 x 10-15min expert fb directly after first practice attempt, focused on assessment domains.	**Late feedback** Same as early fb but at end of second practice session.
Manzone 2014 (verbal comment focused on performance vs verbal comment + comparison to training levels) Canada	Simulated intubation using manikin	**Medical students** UGY1-2	Intervention duration: 1 session Instructional video on intubation. 1-1.5h practice with manikin, with learner in 4 different positions (5 x practice trials in each position). 10 x fb by expert, given directly after practice trials in 2 positions (2 x 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 2 aspects of performance (either done correctly or not) e.g. 'improper use of the laryngoscope'. + individual's progress on task.	**Performance comment compared to others (different training levels)** Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standards at different training levels e.g. 'your performance was at the level of a resident.'
Manzone 2014 (verbal comment on performance vs numerical rating, focused on	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

individual progress)					of hand movements). Plotted on graph to focus on own progress.
Manzone 2014 (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 to 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.
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk 1st session: 2h instruction and practice suturing & tying knots 'until able to do it easily'. 2nd session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30mins familiarisation with equipment. Practice: 1h 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 & given target goals. .	**Simulator output information**
Ozcakar 2009 Turkey	Simulated patient consultation with a simulated patient (communication and history taking skills)	**Medical students** UGY2 62% M	Intervention duration: 2 wk Study conducted during routine university module on clinical skills training. Evaluation: 2 wk after intervention following clinical skills lectures + practice with video recording.	**Video review with expert + expert fb** 1 x videotaping of SP consultation. Directly afterwards, review video with expert plus fb.	**Expert fb** 1 x expert fb directly after SP consultation

Pavo 2016 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 x 2h additional session including training using modified Peyton 4 step approach[b] 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 & depth, hand position, decompression & 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 & depth plus automated audio advice to correct any deviations during CPR.
Rogers 2012 USA	Simulated surgical skill (tying a single 2-handed square knot)	**Medical students** 'surgical clerkship year'	Intervention duration: 1 session Training in knot tying. 2 x practice trials (1 before & 1 after training). Performance evaluation: end of session.	**Expert fb** 1x fb from expert, with specific information on improving subsequent performance, directly after performance.	**Compliment** 1 x general compliment from expert, instead of fb.
Skeff 1983 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 mid- and end of 1 month term.	**Expert peer fb** 1 x 1h session mid-term 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.
Sox 2014 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 x 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 x constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw 1 item evaluation form but not given a copy.

Strandbygaard 2013 Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extra-uterine pregnancy)	**Medical students** UGY 4-6 44% M	Intervention duration: 2 months 1 x 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 3h) for up to 2 months.	**Standardised expert fb with later, additional expert fb if requested by learner** **+ simulator performance score** 1-3 x 10-12min episodes of expert fb involving information on how to perform task components correctly. 1st fb episode provided after first practice trial; learner could ask for up to 2 additional fb episodes (optional) involving same standardised advice + simulator performance score.	**Simulator performance score**
Van de Ridder 2015a (Advances in Health Science Education) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 35% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. Fb from senior medical student with acting experience & 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** 1x fb directly after baseline performance. Fb comment started with global praise followed by the most suitable suggestion for improvement, selected from a list of 4 commonest task errors (e.g. 'You did this well; a tip …')	**Negatively framed fb** 1x episode fb directly after practice trial. Fb comment started with global criticism followed by most appropriate directive advice for improvement, selected from list of 4 commonest task errors. (e.g. 'You did not do this correctly; you should change.')
Van de Ridder 2015b (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 31% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. All fb providers trained for 1h 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 x fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag & introduced as Professor, NT, wearing a white coat).	**Low credibility fb provider** 1 x fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female, informally dressed).

Xeroulis 2007 Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. 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.

Prior experience

Participants were novice to the task in 11 studies (11/20, 55%). Three studies documented prior experience: one study involved attending physicians teaching on ward rounds with a range of experience[30] and two studies documented previous training including CPR[24] and history taking and communication skills in medical students.[43] The remaining six studies that did not report this information. One of these studies[45] involved teaching CPR to first year health professional students across a mix of disciplines, some of whom may have had prior experience with the task. One study[47] involved evaluating professional and clinical skills in first year paediatric residents who are likely to have had relevant training as medical students. In two studies of these studies, participants' baseline performance of junior medical students' surgical knot tying was poor, which suggests no or little prior experience.[35,36] In the last two studies there was no information on prior experience: one assessed a simulated medication consultation by third year pharmacy students[48] and another[42] assessed case presentation skills in third year medical students in their paediatric attachment.

Baseline performance

As before, although randomisation should account for differences in baseline performance, it can be useful information in studies with small samples. There was unequal baseline performance between groups reported in one study[41] and identified from another study's data (obtained from authors)[34]. No statistically significant differences in baseline performance between groups were reported in seven studies [28,30,35,37,38,41,47] and baseline performance was not reported in eleven studies.[22,24,29,36,39,42-46,48]

Research funding

Regarding research funding, one study[24] was loaned a device by Philips as detailed earlier, seven studies received funding from independent institutions,[28,30,37,39,42,44,47] six studies did not

receive any funding[22,35,38,40,41,43] and six studies did not report information on funding.[29,34,36,45,46,48]

Comparison 2: Heterogeneity in feedback interventions

Each study included at least one verbal face-to-face feedback group, in accordance with the inclusion criteria. The alternative feedback interventions were markedly diverse. Table 2 describes the intervention and control conditions for included studies and Figure 5 presents the forest plot and standardised mean differences (SMD), accompanied by the GRADE risk of bias assessment. One additional study[45] 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 odds ratio of 1.25 (95% CI: 0.85-1.84) that favoured the learning conversation. Due to the heterogeneity between studies, we did not pool the outcomes in a meta-analysis.

Once again, the verbal face-to-face feedback interventions differed regarding feedback source, timing and content. Also once again, studies varied across other factors reported to influence performance outcomes,[14,15] including prior experience of participants, instruction and expert demonstration of the task and opportunity to practice.

Some studies investigated straightforward variations in feedback, including differences in frequency (low or high[37]), stage of practice (early or late[38]), different feedback models ('learning conversation' compared with 'feedback sandwich' frameworks[45]), source expertise (expert or peer[22]) and expert feedback compared to compliments.[35,36] Another collection of studies explored the effect of adding expert feedback to other interventions, such as in addition to simulator performance data[24,44] or to written feedback;[30,47] or adding expert review of a participant's performance video to a practice session in which expert feedback was available.[46] One study[34] compared verbal feedback by an expert who had just directly observed the

performance, with written feedback emailed later that day by another expert who watched a video of the performance. Other studies explored more complex phenomena. One study[39] compared two feedback variations in different combinations across four groups. One variation compared an evaluative verbal comment from an expert, to a written numerical performance rating. The second variation involved an individual comparing their performance evaluation to either their own previous attempts (highlighting individual progress) or to expected performance at student, resident or specialist level (comparison with others). Another research group investigated two complex influences in separate studies. One study[41] examined how the credibility of the feedback provider (high or low credibility) influenced learner outcomes. The other study[40] examined the effect of phrasing corrective information in different ways, so in one intervention corrective information was framed within a positive phrase whereas in the other, it was framed within a critical phrase.

Comparison 2: GRADE quality of evidence assessment

Risk of bias analysis

The GRADE risk of bias assessment for the comparison of verbal face-to-face feedback to alternative feedback is presented in Figure 4. In summary, all studies were appropriately randomised; many studies did not provide information on allocation concealment; none could blind participants or research team members due to the face-to-face intervention; most had outcomes assessed by blinded assessors or a machine; most had high proportions of participant completion data; most did not have a prior published protocol but did present expected outcomes.

DISCUSSION

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

Our analysis found that verbal face-to-face feedback substantially enhances health professionals' performance compared to no feedback. The standardised mean difference (SMD) in workplace task performance involving only higher quality studies was 0.7 (95% CI 0.37-1.03; $P<0.0001$) in favour of feedback. With all the included studies, SMD was 1.09 (95% CI 0.59-1.59; $P<0.00001$). Both results are dominated by one study[24] which contained the majority of participants in the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02-0.51). This study involved medical students performing CPR on a manikin. During a brief practice (few minutes), the student who was performing ventilation was responsible for providing the feedback information to the student performing chest compressions, who's performance was being assessed. At the start of each CPR cycle, feedback included performance information, such as compression depth and rate, and advice on how to correct it, if necessary. The short practice period and brief feedback from a peer (as opposed to an expert) who was concurrently performing a task are factors that may have contributed to the relatively small overall benefit reported, when this intervention was compared to practice without feedback (both groups were provided with instruction and expert demonstration).

The consistent positive effects across all included studies supports that the average effect of verbal face-to-face feedback in the health professions is very likely to enhance performance. We assessed the overall quality of the evidence to be moderate (see Figures 2 and 4). The pooled effect size was moderately large at 0.7;[32] this indicates that someone at the 50th centile (i.e. mean performance score) in the feedback group would be at the 76th centile performance score in the no feedback group.[49,50] To give more context regarding the magnitude of effect sizes when any kind of feedback is compared to no-feedback, the meta-analysis by Kluger and

DeNisi[51] reported a pooled SMD of 0.4, although one third of included studies reported a detrimental impact. The largest beneficial effects were seen for interventions that i) included performance information about what had changed since the previous attempt (effect size 0.55), ii) set a specific and challenging goal (effect size 0.51), iii) posed little threat to self-esteem (effect size 0.47) and iv) included information on the correct outcome (effect size 0.43).

The calculated prediction interval for the comparison of verbal face-to-face feedback to no feedback in the higher quality studies was -0.06 to 1.45. This indicates a wide likely range in feedback effect for any individual study, from a very small detrimental impact to a very large beneficial effect on performance, and raises the question regarding whether the learning conditions, performance targets and nature of feedback might combine to create this distributed spectrum of results. Our analysis adds valuable information to previous systematic reviews by Veloski et al[14] and Ivers et al[15] which reported that an audit and feedback process enhanced health professionals' performance.

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

For the second comparison of the effect of verbal face-to-face feedback compared to alternative feedback on performance, there was a diverse range in the alternative feedback interventions and hence we did not conduct a meta-analysis. Beneficial effects were reported by one or two studies each regarding additional expert coaching sessions compared to routine monthly written feedback from attending doctors;[47] expert feedback at the beginning of practice compared to later;[38] additional episodes of feedback from experts;[37,44] additional episodes of feedback involving expert video analysis[46] and expert feedback compared to compliments.[35,36]

Factors that influence performance: feedback and others

The studies we included in the review differed considerably in many aspects that may influence performance. These included feedback components (such as source, timing and content), participants (such as seniority, experience or motivation), performance tasks (such as task complexity, or nature e.g. predominantly cognitive, like ward round teaching or psychomotor, like laparoscopic surgery), contemporaneous teaching, practice time and intervention duration.

Variations in verbal face-to-face feedback interventions: source, timing and content

Focusing on verbal face-to-face feedback interventions, three factors in included study designs have been reported to modify the effect of feedback on performance: feedback source, timing and content. A highly credible source was reported to enhance the effect of feedback and audit in earlier systematic reviews[14,15] and other research.[52-56] In our meta-analysis, all comparisons involved expert feedback, except two.[22,24] In the comparison of verbal face-to-face feedback to alternative feedback, one study directly compared expert feedback to peer feedback[22] but did not find a statistically significant difference (SMD 0.46), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother.

Studies on the timing of feedback have indicated that feedback during practice results in faster initial skill acquisition (as errors are corrected in real-time), particularly for procedural skills, but poorer subsequent performance without feedback, suggesting a reliance on it.[28,52,57] When feedback is provided directly after performance, the learning occurs more slowly but is more enduring, compared to feedback during practice. This is thought to reflect that a person constructs and refines their mental schema for a task as they refine a skill, which they utilise when required to perform the task later. Feedback during performance is thought to risk cognitive overload; when a learner pays attention to the feedback, this detracts from a learner's cognitive processing capacity to work out themselves how to do the task and enhance their

mental schema, which results in reduced subsequent independent performance.[38,58] However, for those with no prior experience or limited capability who are learning a complex task, early feedback may prevent extreme frustration which may lead to giving up.[44,52] Prolonged delay in feedback typically reduces the benefit, as a learner's attention and problem solving has moved on to other things and the memory of the task performance is no longer fresh. The one study included in our review that compared feedback during practice to feedback afterwards did not find a statistically significant difference in performance directly afterwards.[28] However, when performance was evaluated a month later (beyond the scope of our review), the 'feedback after' group performed significantly better.

Regarding the content of feedback, there is evidence that feedback enhances performance when the goal is clear (for example, describing the correct outcome or giving an expert demonstration of the task) and contains advice on how to improve.[10,15,51] Across studies included in our review, the detailed specifications about feedback content were often not clearly reported.

Variations in alternative feedback interventions

The studies included in the comparison of verbal face-to-face feedback to alternative feedback revealed a multitude of feedback variations. These included differences in feedback source (expert or peer, high or low credibility), feedback frequency (once or more frequently), feedback provided at different stages of practice (at the start or after some practice), feedback information (video review, detailed or simple performance evaluations), feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or feedback sandwich), phrasing of feedback (expressing the same corrective information in a positively or negatively couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into the multiple interacting components that influence feedback outcomes.

However, we were not able to clearly identify effective feedback components as lines of research enquiry had not been systematically pursued with sufficient replication and rigour, highlighting the need for further validation studies.

Other influences on performance beyond feedback

Two important influences on performance, in addition to feedback, include teaching and practice. Teaching and expert demonstration particularly assist a learner in the initial phase of skill acquisition, when they need to grasp the component steps involved in a task.[59,60] Practice is essential for mastering a skill however, in our review, practice opportunities varied enormously across studies involving similar tasks. For example, one study on knot tying organised almost daily practice for a month,[29] whereas other studies on knot tying only allowed a single practice.[35,36] In addition, other factors that could influence performance included learners' prior relevant expertise (e.g. first year medical student or a surgical trainee performing a surgical task) and the complexity of the task (knot tying or laparoscopic bariatric surgery).

Effects of feedback beyond performance

This systematic review focused on the effect of feedback on objective performance. However, studies reported additional important outcomes that we did not collate but wish to acknowledge. For example, in one study that compared framing the same corrective information in a phrase that commended or criticised the learner, the authors reported that commending the learner resulted in higher self-efficacy scores.[40] Self-efficacy is important for learning and performance as it is associated with increased motivation and likelihood of achieving a goal.[10,12] Another two studies that compared expert feedback to compliments reported that those learners who received compliments were more satisfied but only those given specific suggestions for improvement by an expert improved their performance.[35,36] This highlights the risk of using 'learner satisfaction' as an indicator of effective feedback. However,

by educating learners on the difference between feedback and compliments, the greater satisfaction related to compliments was eliminated.[36]

Review strengths and limitations

In this systematic review, we have reported for the first time, the benefit of verbal face-to-face feedback, compared to no feedback. The effect of feedback was positive in all the studies included in the meta-analysis, suggesting a consistent effect. Our review also summarised multiple variations in verbal face-to-face feedback interventions. We have identified the need for a systematic research approach to provide robust high quality evidence on the effects of specific feedback factors.

However, there are 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 the data that would allow easy pooling of data, we either calculated an estimate from available data (including reading off graphs) or excluded the study. A checklist by journal reviewers could correct this for future publications and accelerate our ability to synthesise existing data. Most included studies were conducted in a simulated environment, at Kirkpatrick evaluation level two (change in performance), with only a few situated in authentic clinical practice at Kirkpatrick level three (change in performance at work).

Implications for practice and future research

The diverse feedback interventions we collated in this review highlight the substantial variability encompassed by the term 'feedback intervention'. One reason for the variable effect seen may be the many different ways that feedback interventions were designed, studied and assessed. Many of the included studies were 'one-off', involved small numbers of participants and included sources of bias. To advance this field of knowledge, research programs designed to systematically investigate the feedback components required for effective feedback are

needed. This is likely to involve a series of studies designed to isolate one feedback component at a time, with all other key influences on performance standardised, in order to identify, replicate and validate the conditions that are most effective in helping learners to improve, across different contexts.

Summary

Verbal face-to-face feedback in the health professions substantially enhances health professionals' workplace task performance compared to no feedback SMD 0.7 (95% CI: 0.37-1.03; P<0.0001). Future research should focus on systematically analysing components that maximise the effects of feedback.

Declarations of interest: The authors report no declarations of interest.

Author Contributions: CJ and JK designed the review, CJ created the search strategy and undertook the literature searches, CJ and MW undertook study screening, data extraction and quality assessment with oversight from JK, CJ and JK undertook data analysis and interpreted the findings, CJ prepared the manuscript and all authors contributed to the final version.

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

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: The authors report no competing interests.

Patient consent: Not required.

Ethics Approval: Not required.

Data sharing statement: Data access can be requested by contacting the corresponding author.

Figure and Table Legends

Figure 1: PRISMA flow diagram.

Figure 2: Risk of bias assessment and forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 3: Funnel plot of the comparison of the effect of feedback, compared to no feedback, on performance.

Figure 4: Summary of findings table for the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 5: Risk of bias assessment and forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance.

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

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

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. Vickery AW, Lake FR. Teaching on the run tips 10: giving feedback. *Med J Aust* 2005;183(5):267-8. [published Online First: 2005/09/06]

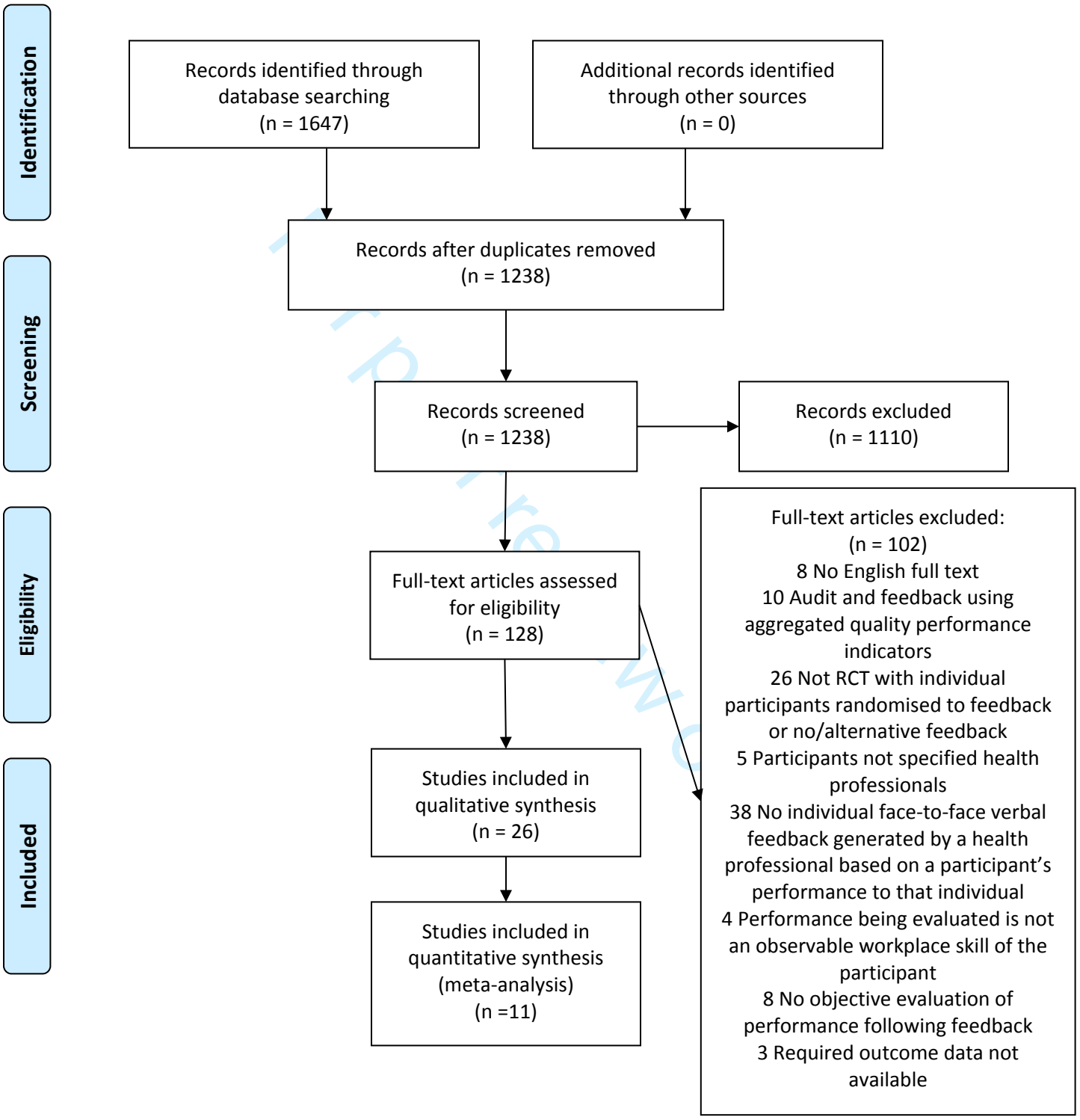
3. Sargeant J, Mann K. Feedback in medical education: skills for improving learner performance. In: Cantillon P, Wood D, eds. ABC of Learning and Teaching in Medicine. 2nd ed. Oxford: Blackwell Publishing Ltd 2010:29-32.
4. Norcini J. Workplace assessment. In: Swanwick T, ed. Understanding Medical Education: Evidence, Theory and Practice. 2nd ed. Oxford: John Wiley and Sons Ltd 2014:279-92.
5. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association* 2009;302(12):1316-26. doi: https://dx.doi.org/10.1001/jama.2009.1365
6. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
7. Ende J. Feedback in clinical medical education. *Journal of the American Medical Association* 1983;250(6):777-81. doi: 10.1001/jama.1983.03340060055026
8. Holmboe ES. Realizing the promise of competency-based medical education. *Academic Medicine* 2015;90(4):411-3. doi: 10.1097/acm.0000000000000515 [published Online First: 2014/10/09]
9. Molloy E, Boud D. Changing conceptions of feedback. In: Boud D. ME, ed. Feedback in higher and professional education London: Routledge 2013:11-33.
10. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol* 2002;57(9):705-17. [published Online First: 2002/09/20]
11. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine* 2015;90(5):609-14. doi: https://dx.doi.org/10.1097/ACM.0000000000000560
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. *Medical Teacher* 2011;33(12):961-73. doi: https://dx.doi.org/10.3109/0142159X.2011.595435
13. 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 Medical Education* 2016;16(1):96. doi: 10.1186/s12909-016-0613-5
14. Veloski J, Boex JR, Grasberger MJ, et al. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
15. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
16. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal* 2008;336(7650):924-26. doi: 10.1136/bmj.39489.470347.AD
17. Ryan R, Hill, S. How to GRADE the quality of the evidence. Cochrane Consumers and Communication Group 2016 [Version 3.0:[Available from: http://cccrg.cochrane.org/author-resources.
18. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List. *Journal of Clinical Epidemiology* 1998;51(12):1235-41. doi: 10.1016/S0895-4356(98)00131-0
19. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *British Medical Journal* 2003;327(7414):557-60.
20. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal* 2011;342:d549.
21. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
22. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058

23. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
24. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
25. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
26. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
27. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
28. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
29. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
30. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
31. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
32. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration 2011 [Available from: www.handbook.cochrane.org.
33. Sedgwick P. Meta-analyses: how to read a funnel plot. *British Medical Journal* 2013;346 doi: 10.1136/bmj.f1342
34. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
35. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
36. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009
37. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5
38. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
39. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397

40. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
41. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
42. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209
43. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
44. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e
45. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
46. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
47. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine* 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44
48. 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(1):322-28.
49. Sullivan GM, Feinn R. Using Effect Size -or Why the P Value Is Not Enough. *Journal of Graduate Medical Education* 2012;4(3):279-82. doi: 10.4300/JGME-D-12-00156.1
50. Coe R. It's the effect size, stupid: What effect size is and why it is important. A paper presented at the Annual Conference of the British Educational Research Association 2002. 2002
51. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
52. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
53. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
54. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
55. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.
56. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
57. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360

58. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
59. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
60. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.

Figure 1: PRISMA flow diagram for systematic review of verbal face-to-face feedback compared to no or alternative feedback.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

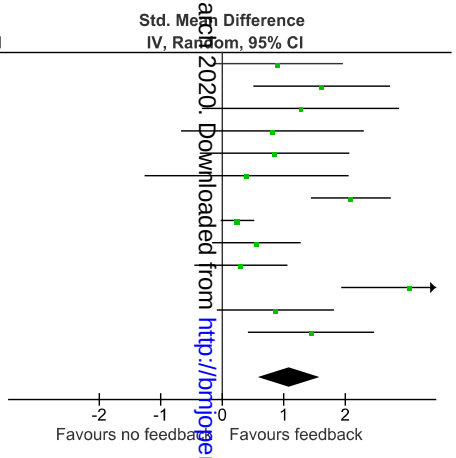
Figure 2: Risk of bias assessment and forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance

a) All included studies

Risk of Bias	A	B	C	D	E	F
Ahlborg 2015						
Bonrath 2015						
Boyle 2011						
Kroft 2017						
O'Connor 2008						
Olms 2016						
Pavo 2016						
Skeff 1983						
Soucisse 2017						
Vafei 2017						
Xeroulis 2007						

	Feedback			No feedback				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.6%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.3%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	5.3%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	5.7%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.9%	0.85 [-0.35, 2.06]
O'Connor 2008	-1.96	1.25	3	-2.5	0.9	3	5.1%	0.40 [-1.25, 2.04]
Olms 2016	-1.1	1.8	30	-7.4	3.8	30	9.6%	2.09 [1.45, 2.73]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	11.0%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	9.2%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	9.1%	0.30 [-0.44, 1.05]
Vafei 2017	17.73	1.09	15	12.13	2.29	15	7.4%	3.04 [1.95, 4.13]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	8.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	7.7%	1.44 [0.43, 2.46]
Total (95% CI)			**253**			**235**	**100.0%**	**1.09 [0.59, 1.59]**

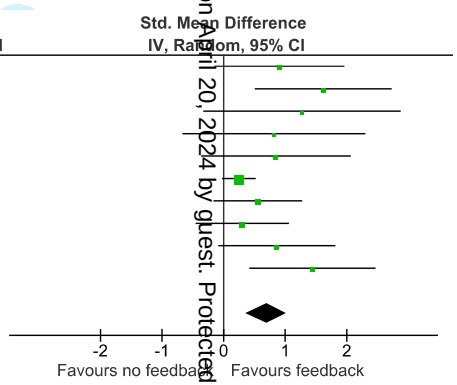
Heterogeneity: Tau² = 0.58; Chi² = 54.26, df = 12 (P < 0.00001); I² = 78%
Test for overall effect: Z = 4.26 (P < 0.0001)



b) Higher quality studies only (sensitivity analysis)

	Feedback			No feedback				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.7%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.1%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	3.8%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	4.4%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.2%	0.85 [-0.35, 2.06]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	27.9%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	13.3%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	12.4%	0.30 [-0.44, 1.05]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	9.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	8.1%	1.44 [0.43, 2.46]
Total (95% CI)			**205**			**187**	**100.0%**	**0.70 [0.37, 1.03]**

Heterogeneity: Tau² = 0.08; Chi² = 13.64, df = 9 (P = 0.14); I² = 34%
Test for overall effect: Z = 4.14 (P < 0.0001)



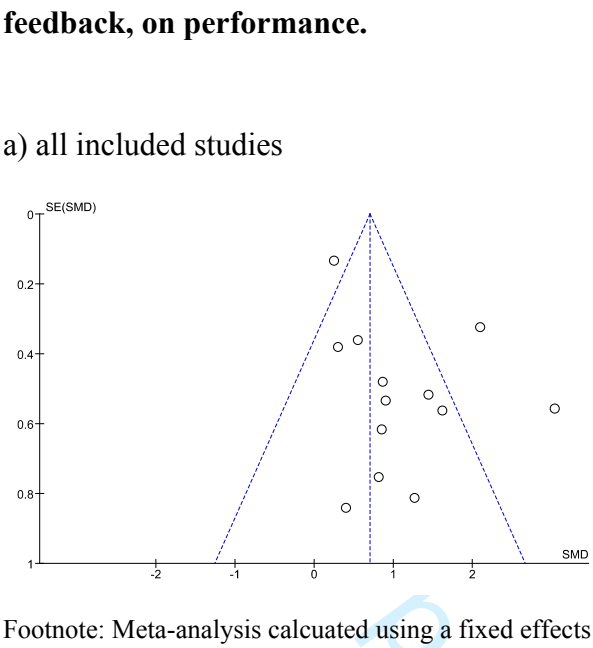
Risk of bias legend ☐ low risk ☐ unclear risk ■ high risk
A: Random sequence generation (selection bias)
B: Allocation concealment (selection bias)
C: Blinding of participants and research staff (performance bias)
D: Blinding of outcome assessment (detection bias)
E: Incomplete outcome data (attrition bias)
F: Selective reporting (reporting bias)

Footnotes
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

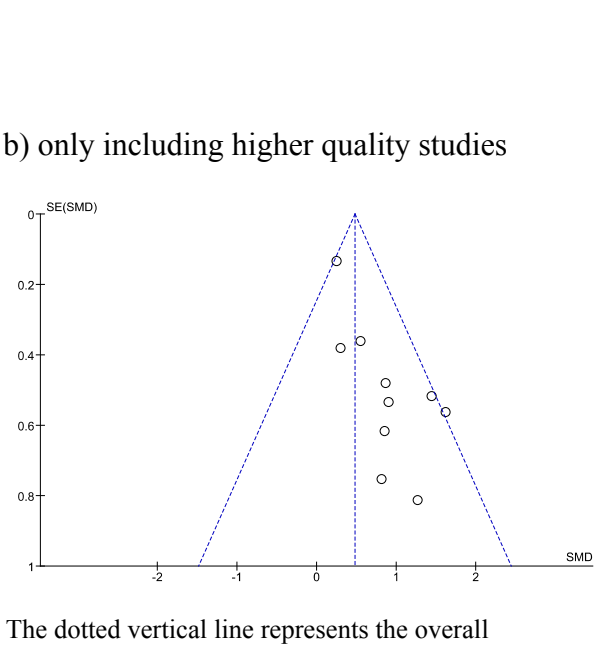
Abbreviations
SD = standard deviation
CI = confidence interval
SMD = standardised mean difference

Figure 3: Funnel plot of the comparison of the effect of feedback, compared to no feedback, on performance.

a) all included studies



b) only including higher quality studies



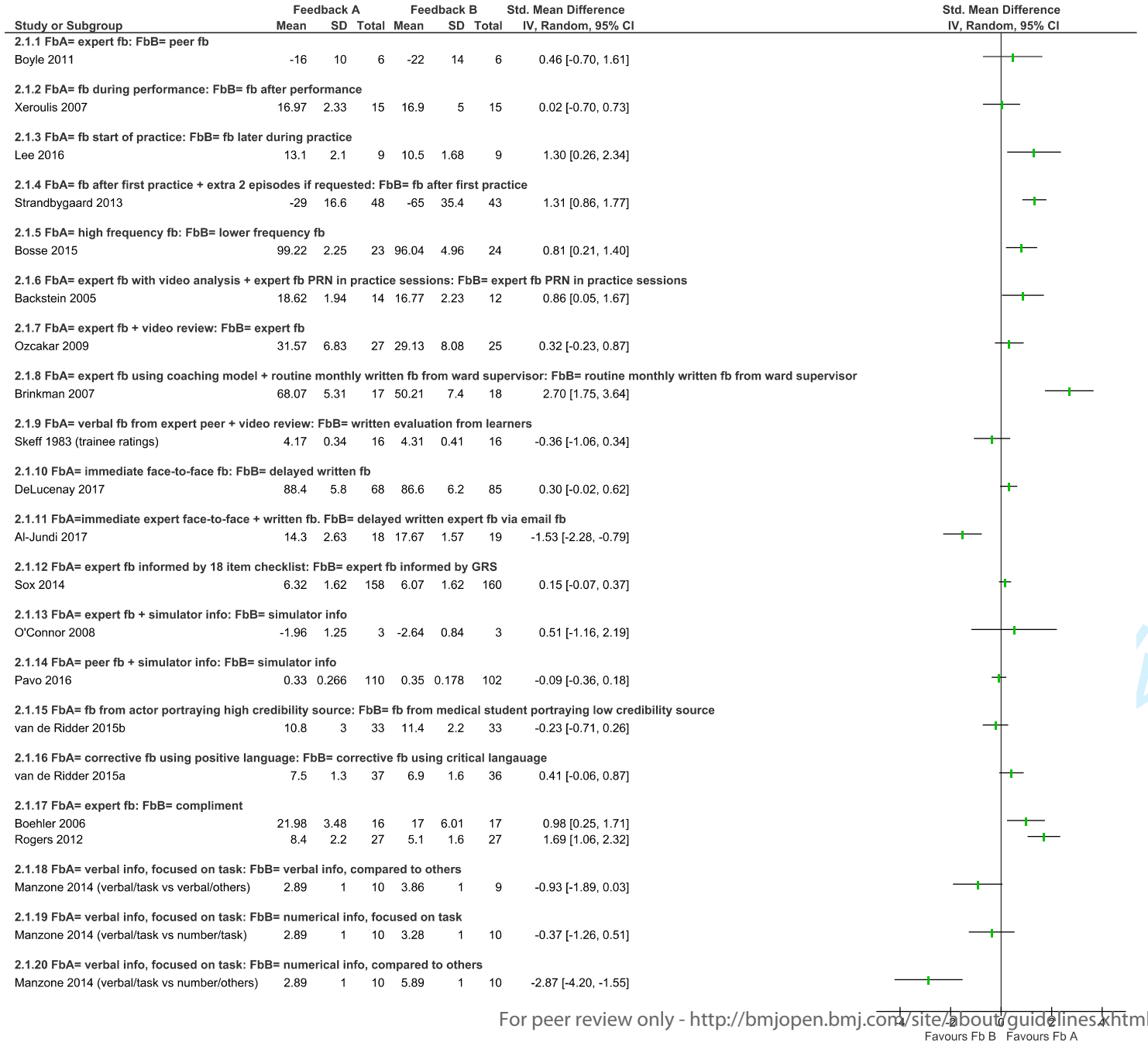
Footnote: Meta-analysis calcuated using a fixed effects model. The dotted vertical line represents the overall effect estimate and the dotted slanted lines represent the 95% confidence interval lines.

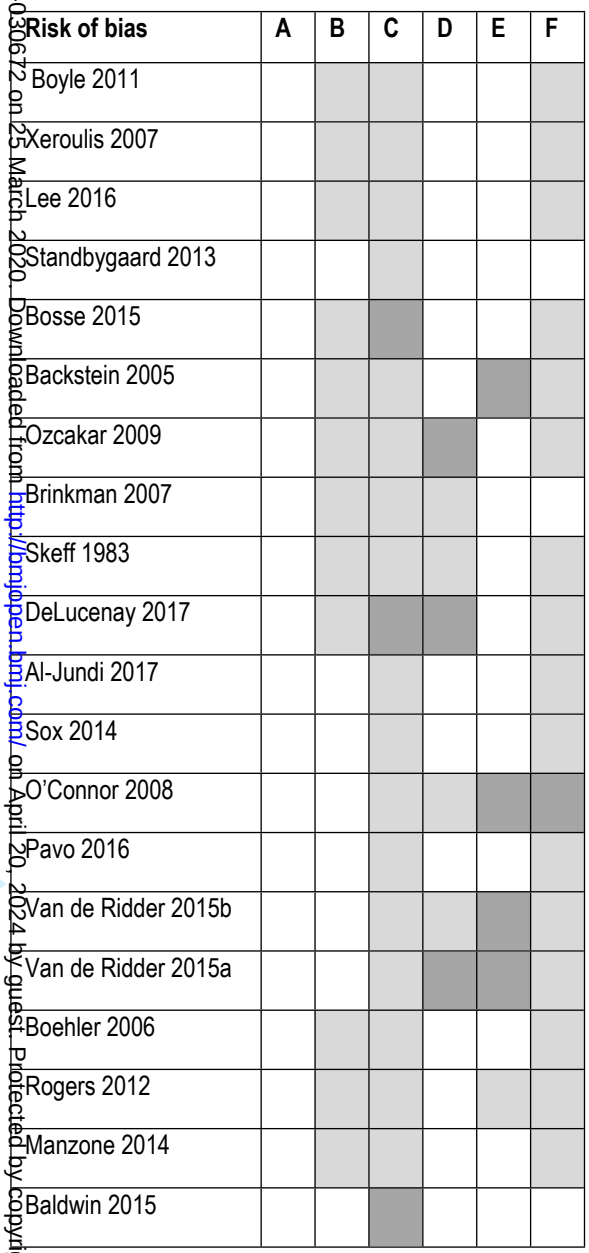
Abbreviations: SE = standard error; SMD = standardised mean difference

Figure 4: Summary of findings table for the effect of verbal face-to-face feedback, compared to no feedback, on performance.

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**	**Quality of evidence (GRADE)**	**Comments**
Objective assessment of observed performance	The mean score in the intervention group was 0.7 standard deviations (0.37-1.03) higher than mean scores for the control group	Number of participants 392 (8 studies)	⊕⊕⊕O Moderate *Due to risk of bias, inconsistency and publication bias*	A SMD of 0.7 indicates a substantial improvement in performance
CI = Confidence interval; SMD= standardised mean difference				
GRADE Working Group grades of evidence ⊕⊕⊕⊕ High: We are very confident that the true effect lies close to that of the estimate of the effect. ⊕⊕⊕O Moderate: We are moderately confident in the effect estimate, the true effect is likely to be close to this but there is a possibility that it is substantially different ⊕⊕OO Low: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect ⊕OOO Very Low: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.				

Figure 5: Risk of bias assessment and forest plot for the effect of verbal face-to-face feedback (Feedback A) , compared to alternative feedback (Feedback B), on performance





Abbreviations

Fb= feedback; GRS= global rating scale; info= information; PRN= 'as required'
SD = standard deviation; CI=confidence interval

Footnotes:
[1]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 standard deviation) from author
Rogers 2012: additional information (standard deviation) 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
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

Risk of bias legend ☐ low risk ▒ unclear risk ■ high risk

A: Random sequence generation (selection bias)
B: Allocation concealment (selection bias)
C: Blinding of participants and research staff (performance bias)
D: Blinding of outcome assessment (detection bias)
E: Incomplete outcome data (attrition bias)
F: Selective reporting (reporting bias)

Database(s): Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:

#	Searches	Results
1	*Feedback/	6031
2	Feedback, Psychological/	3311
3	Formative Feedback/	467
4	(feedback adj3 (effective or formative or constructive or quality or clinical or performance)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4860
5	1 or 2 or 3 or 4	13233
6	exp Health Personnel/	470058
7	exp Health Occupations/	1648689
8	exp Dentistry/	386159
9	exp Social Work/	17331
10	exp Psychology/	66579
11	Occupational Therapy/	13213
12	Radiotherapy/	42757
13	Radiography/	334082
14	Mentors/	9949
15	exp Students, Health Occupations/	60760
16	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2722708
17	clinician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	171551
18	(health* adj2 (staff or personnel or faculty or provider* or worker* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	303659
19	doctor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104240
20	physician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	495037
21	(medical adj3 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	133207
22	general practitioner*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45015
23	(general pract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45488
24	(family adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21549
25	(primary care adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10661
26	(primary health* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2156
27	(registrar or registrars or senior house officer* or resident or residents or hospital medical officer* or intern or interns or house officer*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	129452
28	dentist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	115221
29	(dent* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21015
30	nurs*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	629304
31	(midwife or midwives).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	18155

32	(midwife* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	965
33	(allied health adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	12936
34	physiotherapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5348
35	physical therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5123
36	(physiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	731
37	(physical therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	647
38	occupational therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4500
39	(occupational therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1058
40	speech therap*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7630
41	(speech therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	22
42	speech language therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	136
43	(speech language therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	14
44	speech pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	656
45	(speech pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	31
46	speech language pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1561
47	(speech language pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	105
48	dietician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1352
49	(dietetic* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	603
50	podiatrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	626
51	(podiatry adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	19
52	chiropodist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	119
53	(chiropody adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2
54	social worker*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8079
55	(social work* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8517

56	psychologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	11986
57	(psychology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1532
58	(osteopath* or osteopathic physician*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	6480
59	(osteopath* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	296
60	chiropractor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1148
61	(chiropract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	241
62	pharmacist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	27316
63	(pharmac* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7184
64	optometrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1764
65	(optometr* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	290
66	(Radiographer* or radiological technologist* or radiation therapist* or radiotherapist* or radiation therapy technologist*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2449
67	(radiograph* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	267
68	(radiation therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	53
69	(radiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	100
70	(supervisor* or tutor* or trainer* or educator* or teacher* or mentor* or preceptor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104110
71	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70	1846616
72	16 or 71	3625236
73	exp Education, Professional/	284462
74	exp Educational Measurement/	137138
75	exp Professional Practice/	247602
76	exp Simulation Training/	6239
77	(effect* or evaluat* or outcome* or assess* or measur*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10754772
78	73 or 74 or 75 or 76 or 77	11134144
79	randomized controlled trial.pt.	505181
80	controlled clinical trial.pt.	100406
81	randomized.ab.	391590
82	randomly.ab.	266043
83	systematic review.ab,ti.	85419
84	79 or 80 or 81 or 82 or 83	966230
85	5 and 72 and 78 and 84	821
86	limit 85 to (english language and humans)	809

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	P1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	P2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	P4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	P5-6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	P6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	P6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	P6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	P6-7 & Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	P6-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	P6-8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	P7, Tables 1 & 2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	P7-8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	P9

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	P9-10

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	P7-8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	P9-10
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	P10 & Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	P10-12 & 17-20 Tables 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	P14-16, & 20 & Figures 2 & 5
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures 2 & 5
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	P9 & Figures 2-4
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	P9-11, 20 Figures 2 & 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	P9-10, Figures 2,4 & 5
DISCUSSION			

PRISMA 2009 Checklist

Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	P21-22
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	P26
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	P26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	P27

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030672.R1
Article Type:	Original research
Date Submitted by the Author:	01-Nov-2019
Complete List of Authors:	Johnson, Christina; Monash Health, Monash Doctors Education ; University of Melbourne, Department of Medical Education, Melbourne Medical School Weerasuria, Mihiri; Monash Health Keating, Jennifer; Monash University, Department of Physiotherapy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	MEDICAL EDUCATION & TRAINING, feedback, health professions education, formative feedback

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Short title: Systematic review and meta-analysis on feedback

Christina E. Johnson MBChB, MHPE, FRACP

ORCID ID: 0000-0002-4209-8419

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, Australia.

Mihiri P. Weerasuria MBBS (Hons)

ORCID ID: 0000-0001-6897-3982.

Monash Health, Melbourne, Australia.

Jennifer L. Keating PhD, GradDipManip PT, BAppSci (PT)

ORCID ID: 0000-0003-3161-4964

Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia.

Correspondence should be addressed to Christina Johnson, Monash Doctors Education, McCulloch House, Monash Health, Clayton Road, Clayton, Victoria 3168, Australia. Telephone: 613-9594 3743. Email: 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 1st 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 to no feedback or alternative feedback, and available as full text publications in English. The certainty of evidence was assessed using the GRADE approach. For feedback compared to no feedback, outcome data from included studies was pooled using a random-effects model.

Results: In total, 26 trials met the inclusion criteria, involving 2307 participants. For verbal face-to-face feedback compared to 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 0.37-1.03; P<0.001) in favour of feedback. The calculated SMD prediction interval was -0.06 to 1.46. For feedback compared to alternative feedback, studies could not be pooled due to substantial intervention and design

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: We are moderately confident that verbal face-to-face feedback in the health professions substantially enhances workplace task performance, compared to no feedback. We found a lack of high-quality trials that clearly reported key components likely to influence performance. To build a robust evidence-base, and identify best practice in feedback, a standardised approach to investigations and reporting is required.

PROSPERO 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 to 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.

• For the meta-analysis, not all studies that met the inclusion criteria clearly reported the data required for pooling, so for some studies data was obtained from the author or estimated by calculating it using available data or reading off graphs; otherwise, the study was excluded.

Keywords: feedback, effective feedback, formative feedback, systematic review, meta-analysis, health professions education

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 (for example, from a simulator or within an online learning module). The unique potential benefit of face-to-face verbal feedback is the opportunity for i) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to co-construct insights and strategies for improvement and ii) inter-personal 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 said, "Without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all".[17] 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 which investigated the impact of audit and feedback. In 2006, Veloski et al 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.[18] 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 commonest data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers et al updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared to usual care.[19] 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 in favour of feedback was small at 4.3% (IQR 0.5-16%), the 75th centile of 16% suggested 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 to just performance information (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 e.g. 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) Registration ID CRD42017081796.

Eligibility criteria for considering studies for this review

We included randomised controlled trials in which individual health professionals were randomised to feedback, compared to 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 pre-determined 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 i) verbal face-to-face feedback compared with no feedback and ii) verbal face-to-face feedback compared to alternative feedback.

Performance following feedback interventions had to be objectively assessed. To isolate the effects of feedback, 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 Appendix 1 for the full search strategy for OvidMEDLINE). 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 1st February 2019. We also searched the reference lists of systematic reviews and included studies.

Data collection and analysis

Selection of studies

One review author (CJ or MW) screened titles to exclude clearly irrelevant reports. Two authors (CJ and MW) independently screened remaining abstracts to identify potential eligible studies, then independently assessed the full text. Decisions were compared using Covidence (on-line 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 (JK).

Data extraction and management

One review author (CJ) used a pre-piloted standardised form to extract data from included studies and another author (MW or JK) 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.

Assessment of risk of bias in included studies

The risk of bias was independently assessed by two authors (CJ and JK) 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 (for example means (standard deviations SD) or medians (interquartile range IQR)). The effect was quantified using the standardised mean difference to combine studies measuring the same outcome (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).[21]

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 results of the random-effects meta-analysis was

presented as the standardised mean difference (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.[24]

We rated the overall certainty of evidence for the outcome using the GRADE approach (Chapter 12, Cochrane Handbook),[21] which considers within-study risk of bias, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias.[21] 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.

[Figure 1 here]

Comparison 1: The effect of verbal face-to-face feedback, compared to 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 to no feedback on the objective evaluation of a workplace task. Seven (64%) reports were published in the last five years since 2014. The studies were conducted in Europe (4),[25-28] Canada (4),[29-32] the USA (2)[33, 34] and Asia (1).[35]

There were 488 participants, including 196/366 (53.6%) males from seven studies that reported gender data.[25, 27-30, 33, 35] Participants included 290 (60%) medical students in four studies,[25, 28, 32, 33] 60 (12%) dental students in one study[27] and 138 (28%) doctors in four studies.[26, 29-31, 34, 35] 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 [29, 30, 34, 35] and a simulation environment in seven studies (7/11, 64%).[25-28, 31-33] 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 online supplementary material for more details - Appendix 2).

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

Author Year Country	Task	Participants Health Profession % Male	Teaching and Practice Same for Feedback Intervention and Control groups	Feedback Intervention										
				Additional information	Source		Timing			Content				
					Subject Expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info[a]	Performance video	Written performance info
Ahlborg 2015 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 x practice trials. Performance evaluation: end of session.	2 x fb episodes. Fb given by expert i) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus ii) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓		
Bonrath 2015 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.) x 25 min fb episodes. Fb given by expert using specific coaching model[b] + video review of learner operating + video exemplars of good/poor technique. Effectiveness of strategies reviewed at subsequent session.	✓				✓	✓	✓		✓	

Boyle 2011 (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 x practice trials Performance evaluation: end of session	5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' + simulator output information.	✓			✓		?	?	✓		
Boyle 2011 (peer fb)	Same as above	Same as above	Same as above	5 x fb episodes. Peer discussed simulator output, any task errors & teaching instructions given at start.		✓		✓		✓	✓	✓		
Kroft 2017 Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy)	Doctors training in O&G PGY2-6 33% M	Intervention duration: 1 x 15min practice using laparoscopic salpingectomy module on VR surgical simulator within 1h of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1 x fb episode from expert directly after VR simulator practice. Fb 'standardized and given in an evidence based fashion to optimise effectiveness' & included '3 constructive recommendations based on performance'.	✓			✓		✓	✓			
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	Medical students UGY1-2 44% M	Intervention duration: 4 wk. 2h instruction + practice suturing & knot tying until able to do it easily. Then instruction on laparoscopic surgery + expert demonstration video of task tying, followed by 30 mins familiarisation with equipment. Practice: 1h 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.	✓		✓	✓		✓	✓	✓		
Olms 2016 Germany	Simulated colour matching teeth	Dental students UGY3	Intervention duration: 1 session Study conducted during 10 wk routine university module on matching tooth shades involving	1 x expert fb session. Fb included correct response + explanation with expert	✓			✓		✓	✓			✓

			variety of teaching + practice opportunities. Performance evaluation: 2 wks after intervention (within one university module).	demonstration if needed + written copy of evaluation. Expert trained in fb.									
Pavo 2016 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 x 2h additional training session: instructional video + training using modified Peyton 4 step approach.[c] Brief practice (few mins) 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 & depth, hand position, decompression & hands-off time. Instructional video for intervention group had demonstrated this.		✓	✓			✓	✓			
Skeff 1983 USA	Clinical teaching skills during ward round	Physicians	Intervention duration:1 session in the middle of 4wk ward duty. At mid & 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 wk later, at end of ward duty	1 x 60 min fb discussion with peer, including video review, trainee ratings & self-assessment to enable physician to identify strengths & devise solutions to problems.	✓				✓	✓	✓		✓	✓
Soucisse 2017 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 wk later (ongoing clinical work as a surgical resident).	1 x 30min expert fb sometime after baseline performance with video review of baseline performance + coaching using 'GROW' model[d] including 2-3 suggestions for improvement + expert demonstration followed by learner demonstration of	✓				✓	✓	✓		✓	

				desired improvements, as required + action plan.										
Vafaei 2017 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 x 5min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓				
Xeroulis 2007 (fb after) Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	Medical students UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. Performance evaluation: end of session.	Expert fb as needed (expert or learner initiated), after practice trials, involving constructive ways to improve + expert demonstration.	✓			✓		✓	✓			
Xeroulis 2007 (fb during)	Same as above	Same as above	Same as above	Same as above except expert fb during practice trials.	✓		✓			✓	✓			

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',[29, 31] six studies were rated 'unclear'[25, 26, 28, 30, 32, 34] and three studies were 'high risk'.[27, 33, 35]

[Figures 2 and 3 here]

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.

[Figure 4 here]

Meta-analysis

A meta-analysis of verbal face-to-face feedback compared to no feedback included 13 comparisons from the 11 studies, involving 488 participants. Two studies reported data that each enabled two comparisons: in one study, feedback provided *during* practice in one group and *directly after* practice in another was compared to no feedback;[32] in another study, feedback provided by an *expert* in one group and by a *peer* in another group[26] was compared to no feedback. In the meta-analysis, numbers for the control group for these studies were halved to retain sample independence.[21]

The meta-analysis of the effect of verbal face-to-face feedback compared to no feedback on workplace task performance found a standardised mean difference of 1.09 (95% confidence interval (CI) 0.59-1.59; $P<0.001$) using a random-effects model. The forest plot is presented in Figure 5a.

[Figure 5 here].

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 ten comparisons.[25, 26, 28-32, 34] The standardised mean difference was 0.7 (95% CI 0.37-1.03; 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 moderate, using the GRADE approach. Figure 6 displays the Summary of Findings table.

[Figure 6 here]

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

Included studies

Table 2 presents the characteristics of included studies in the comparison of verbal face-to-face feedback compared to alternative feedback. Twenty studies (22 comparisons) were included in this analysis and involved verbal, face-to-face feedback compared to alternative feedback. Nine studies (9/20, 45%) were published in the last five 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%) males from 13 studies that reported gender data.[28, 33, 36-46] Included studies involved students (medical, mixed health professions and pharmacy) (1869, 95%) in 16 studies,[28, 32, 33, 36-39, 41-49] and doctors (105, 5%) in four studies.[26, 34, 40, 50] All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[34, 40] Three studies evaluated performance in a clinical practice setting (involving teaching skills,[34] professional and communication skills[40] and oral

case presentations[49]) and the remaining 17 assessed performance in a simulated environment (surgical procedures, nasogastric tube insertion, intubation, hearing test, pharmacy consultation or CPR).[26, 28, 32, 33, 36-39, 41-48, 50] (See online supplementary material for more details - Appendix 2).

Figure 7 presents the forest plot and standardised mean differences (SMD). One additional study[37] 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 odds ratio of 1.25 (95% CI: 0.85-1.84) that favoured the learning conversation. The feedback comparisons were markedly diverse, so we did not pool outcomes in meta-analysis.

[Figure 7 here]

Risk of bias

In summarising the risk of bias across domains within each study, two studies were rated as low risk,[37, 44] seven studies were rated as 'high risk',[33, 41, 45, 46, 48, 50] and the remaining studies were rated as 'unclear'. (See the risk of bias summary in Figure 3).

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

Article First author Year Country	**Task**	**Participants:** Health profession Gender: % Men	**Common to interventions A + B**	**Intervention A** All verbal face-to-face feedback to an individual health professional	**Intervention B**
Al-Jundi[a] 2017 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 x 10 mins for baseline performance. Performance evaluation: 2 days later	**Immediate face-to-face + written expert feedback** 1 x 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 x 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.
Backstein 2005 Canada	Simulated surgical procedure using a bench top model (vascular anastomosis)	**Doctors in surgical training** PGY1	Intervention duration:4 wk Lecture on surgical procedure. 3 x 2h 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 wk 4	**Review of performance video with expert fb** **+ practice sessions with expert fb available** 3 x weekly videotaping of surgical procedure, with expert feedback available during task, followed by up to 15min review of video with expert fb	**Practice sessions with expert fb available**
Baldwin 2015 England	Simulated BLS	**Health profession students** medical (58%), physiotherapy (12%), pharmacy (10%), nursing	Intervention duration: 4 wk Instruction and practice with manikin 3 x 2.5h weekly. Fb provided directly after performance by senior peer instructor. Instructor accredited in	**'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[b], with final summary.	**'Feedback sandwich model'** Fb involved a point for improvement in between 2 points of praise.

		(10%), dentistry (10%) UGY1 33% M	BLS + trained to provide fb. Fb provider compliance monitored. Performance evaluation: in wk 4		
Boehler 2006 USA	Simulated surgical skill using a bench top model (tying a 2-handed square knot)	**Medical students** UGY2-3 52% M	Intervention duration: 1 session Instruction in knot tying from surgeon. 1 x baseline performance. Performance evaluation: end of session.	**Expert feedback** 1 x episode of fb from expert surgeon, directly after performance, describing 1-2 specific ways to improve performance.	**Compliment** 1 x pre-scripted general compliment e.g. 'great job!'
Bosse 2015 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 4 step procedural training method[c] 6 x practice trials. Fb 'positively worded', focused on effect of actions, given directly after performance by senior peer instructors, trained in procedure & fb. Performance evaluation: end of session.	**High frequency fb** 6 x episodes of fb, given after each practice trial.	**Low frequency practice** 2 x episodes of fb, given after first and last practice trial.
Boyle 2011 Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty + stent)	**Doctors training in surgery** PGY4+	Intervention duration: 1 session Teaching & expert demonstration. Fb providers had simulator training. 5 x practice trials (each maximum 40min). Performance evaluation: end of session.	**Expert fb** 5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	**Peer fb** 5 x fb episodes. Peer discussed simulator output, any task errors & the teaching instructions given at start of session.
Brinkman 2007 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 feedback 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.	**Coaching session + routine feedback as part of clinical training** 1 x 30min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses & parents. Used a coaching model[d] focused on assisting learner to understand information, design goals and improvement strategies. Fb given	**Routine feedback as part of clinical training** Performance ratings from nurses and patients not seen.

			Performance evaluation: 5 months after start of clinical attachment.	by paediatricians trained in coaching model.	
DeLucenay 2017 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 5min fb on communication skills. Performance evaluation: last 4 SP consultations.	**Immediate face-to-face fb** 4 x 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 x videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.
Lee 2016 Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis)	**Medical students** UGY3-4 78% M	Intervention duration: 3 wk Instruction and expert demonstration of procedure, followed by 3 x weekly 30min practice sessions. Performance evaluation: end of 3rd session.	**Early feedback** 1 x 10-15min expert fb directly after first practice attempt, focused on assessment domains.	**Late feedback** Same as early fb but at end of second practice session.
Manzone 2014 (verbal comment focused on performance vs verbal comment + comparison to training levels) Canada	Simulated intubation using manikin	**Medical students** UGY1-2	Intervention duration: 1 session Instructional video on intubation. 1-1.5h practice with manikin, with learner in 4 different positions (5 x practice trials in each position). 10 x fb by expert, given directly after practice trials in 2 positions (2 x 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 2 aspects of performance (either done correctly or not) e.g. 'improper use of the laryngoscope'. + individual's progress on task.	**Performance comment compared to others (different training levels)** Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standards at different training levels e.g. 'your performance was at the level of a resident.'
Manzone 2014 (verbal comment on performance vs numerical rating, focused on	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

individual progress)					of hand movements). Plotted on graph to focus on own progress.
Manzone 2014 (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 to 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.
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk 1st session: 2h instruction and practice suturing & tying knots 'until able to do it easily'. 2nd session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30mins familiarisation with equipment. Practice: 1h 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 & given target goals. .	**Simulator output information**
Ozcakar 2009 Turkey	Simulated patient consultation with a simulated patient (communication and history taking skills)	**Medical students** UGY2 62% M	Intervention duration: 2 wk Study conducted during routine university module on clinical skills training. Evaluation: 2 wk after intervention following clinical skills lectures + practice with video recording.	**Video review with expert + expert fb** 1 x videotaping of SP consultation. Directly afterwards, review video with expert plus fb.	**Expert fb** 1 x expert fb directly after SP consultation

Pavo 2016 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 x 2h additional session including training using modified Peyton 4 step approach[b] 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 & depth, hand position, decompression & 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 & depth plus automated audio advice to correct any deviations during CPR.
Rogers 2012 USA	Simulated surgical skill (tying a single 2-handed square knot)	**Medical students** 'surgical clerkship year'	Intervention duration: 1 session Training in knot tying. 2 x practice trials (1 before & 1 after training). Performance evaluation: end of session.	**Expert fb** 1x fb from expert, with specific information on improving subsequent performance, directly after performance.	**Compliment** 1 x general compliment from expert, instead of fb.
Skeff 1983 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 mid- and end of 1 month term.	**Expert peer fb** 1 x 1h session mid-term 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.
Sox 2014 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 x 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 x constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw 1 item evaluation form but not given a copy.

Strandbygaard 2013 Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extra-uterine pregnancy)	**Medical students** UGY 4-6 44% M	Intervention duration: 2 months 1 x 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 3h) for up to 2 months.	**Standardised expert fb with later, additional expert fb if requested by learner** **+ simulator performance score** 1-3 x 10-12min episodes of expert fb involving information on how to perform task components correctly. 1st fb episode provided after first practice trial; learner could ask for up to 2 additional fb episodes (optional) involving same standardised advice + simulator performance score.	**Simulator performance score**
Van de Ridder 2015a (Advances in Health Science Education) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 35% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. Fb from senior medical student with acting experience & 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** 1x fb directly after baseline performance. Fb comment started with global praise followed by the most suitable suggestion for improvement, selected from a list of 4 commonest task errors (e.g. 'You did this well; a tip …')	**Negatively framed fb** 1x episode fb directly after practice trial. Fb comment started with global criticism followed by most appropriate directive advice for improvement, selected from list of 4 commonest task errors. (e.g. 'You did not do this correctly; you should change.')
Van de Ridder 2015b (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 31% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. All fb providers trained for 1h 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 x fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag & introduced as Professor, NT, wearing a white coat).	**Low credibility fb provider** 1 x fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female, informally dressed).

Xeroulis 2007 Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. 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.

DISCUSSION

Comparison 1: The effect of verbal face-to-face feedback on performance, compared to no feedback.

Our analysis found that verbal face-to-face feedback substantially enhances health professionals' performance compared to no feedback, with SMD 0.7 (95% CI 0.37-1.03; *P*<0.001) from eight studies involving 392 health professionals, after excluding studies at high risk of bias. We are moderately certain about the evidence for this outcome. To our knowledge, this is the first report to substantiate the widely held view that feedback enhances performance and to estimate the benefit. Included studies involved health professional students and clinicians (mainly medical) performing a range of workplace tasks, particularly surgical and most commonly in a simulated environment. Both results are dominated by one study[28], evaluating effective compressions during CPR, which contributed the largest number of participants from a single study to the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02-0.51). Several factors may have contributed to the relatively small overall benefit from this feedback intervention compared to many of the other included studies. These include a short practice period with feedback from a peer (as opposed to an expert) who was concurrently performing a different task (the student performing ventilation provided advice on correcting compressions to the student performing compressions).

The consistent positive effects across all included studies supports that the average effect of verbal face-to-face feedback in the health professions is very likely to enhance performance. The pooled effect size was moderately large at 0.7;[21] this indicates that someone at the 50th centile (i.e. mean performance score) in the feedback group would be at the 76th centile performance score in the no feedback group.[51, 52] In comparison a meta-analysis by Kluger and DeNisi[53] which analysed feedback (any type) compared to no feedback, reported a pooled SMD

of 0.4 (notably one third of included studies reported a detrimental impact). To provide more context regarding effect sizes for feedback, the largest beneficial effects of feedback reported in Kluger and DeNisi's meta-analysis were i) effect size 0.55 when feedback included information on any changes since the previous attempt, ii) effect size 0.51 when a specific and challenging goal was set, iii) effect size 0.47 when feedback posed little threat to self-esteem and iv) effect size 0.43 when feedback included information on the correct outcome.

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 likely range in feedback effect for any individual study, from a very small detrimental impact to a very large beneficial effect on performance. This aligns with previous meta-analyses within the health professions and beyond which have reported diverse impacts from different feedback interventions.[14, 18, 19]

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

For the second comparison of the effect of verbal face-to-face feedback compared to 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 additional expert coaching sessions compared to routine monthly written feedback from supervisors;[40] expert feedback early in a practice period compared to later;[42] additional episodes of feedback from experts;[39, 44] additional episodes of feedback involving expert video analysis[50] and expert feedback compared to compliments.[38, 47]

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

Our review focused on the effect of feedback on performance. The studies assembled in this review illustrate the wide variety of possible elements in a feedback intervention. Starting with verbal face-to-face feedback interventions in included studies, important differences were seen between studies, involving feedback content, source and timing. Detailed specifications about feedback content were often not clearly reported, which suggests that researchers may not have realised the importance of this factor. In particular, previous research has identified that feedback is more effective when it makes the goal clear (for example, describing correct performance or providing an expert demonstration of the task) and advice on how to improve.[10, 19, 53] Feedback was more often provided by experts than peers. One small study directly compared expert feedback to peer feedback [26] for novices learning a surgical task using a visual reality simulator. It did not find a statistically significant difference (SMD 0.46, 95% CI -0.7-1.61), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother. In earlier systematic reviews[18, 19] and other research,[54-58] feedback from a highly credible source (expert feedback) has been reported to be more effective. In our review, the timing of feedback also varied; it was provided while the learner undertook the task, immediately afterwards or some time afterwards. One small study, in which novices learnt to suture, compared feedback during the task to feedback immediately after each attempt. It did not find a statistically significant difference in performance after one hour of practice but did a month later (beyond the scope of our review), in favour of feedback immediately after practice.[32] In another study, in which students practised simulated laparoscopic surgery, the effect of additional expert feedback was compared with performance information provided by the simulator alone. The authors reported more participants in the 'simulator feedback only' stopped practising. Previous research has noted that for novices

learning a complex task, early feedback may prevent extreme frustration and giving up.[54] Feedback during task performance results in faster initial skill acquisition compared to feedback after task performance, particularly for procedural skills, as errors are corrected in real-time, but poorer subsequent independent performance.[54, 59] It is thought that a learner develops a mental schema depicting how to do the task, as they work it out during practice attempts and this is utilised for subsequent performances.[60, 61] However feedback during task performance appears to interfere with this process, possibly due to cognitive overload.[42, 62]

In the second analysis verbal face-to-face feedback was compared to a multitude of feedback variations. In addition to feedback (source, frequency, timing) and content, there were differences across feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or 'feedback sandwich'), phrasing of feedback (expressing the same corrective information in a positively or negatively couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into components that might influence feedback outcomes. However, there was insufficient evidence to clearly identify and recommend effective feedback components, as lines of research enquiry had not been systematically pursued with sufficient replication and rigour.

Influences on performance due to factors beyond feedback

Performance improvement is not solely related to feedback. In our review, other important factors influencing performance also varied between studies. Firstly, teaching and expert demonstration were common (but not standard) and the amount and type varied across studies.[63, 64] Practice opportunities also differed enormously across included studies, even those involving similar tasks. In addition, learners' prior relevant expertise (e.g. first year medical students or surgical trainees learning a surgical task) and the complexity of the task

(knot tying or laparoscopic bariatric surgery) varied. Previous research has shown that teaching and expert demonstration assist a learner to improve, particularly in the initial phases of skill acquisition[63, 64] and practise is essential for mastering any skill.[15, 65] In addition, learners who are motivated to learn a challenging but achievable skill are most likely to improve their performance, according to 'goal setting' and 'self-determination' theories.[10, 66]

Review strengths and limitations

In this systematic review, we have reported for the first time, the benefit of verbal face-to-face feedback, compared to no feedback. The effect of feedback was positive in all the studies included in the meta-analysis, suggesting a consistent direction of effect. Our review also summarised and categorised multiple variations in feedback interventions (both verbal face-to-face and alternative feedback) and described other factors that have been reported to influence performance in previous research; this clarifies key parameters that need to be considered in future research into feedback.

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 the 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).[67]

Implications for practice and future research

Our review fills an important evidence-gap and supports the beneficial impact of verbal face-to-face feedback on health professionals' task performance, compared to no feedback. Our review also highlights the substantial variability encompassed by the term 'feedback intervention'. By analysing included studies based on key parameters known to influence performance, our review may assist future researchers to refine their methodology and enrich the value of their results by considering these factors. Many of the included studies were 'one-off', involved small numbers of participants and included sources of bias. To advance this field of knowledge, research programs designed to systematically investigate the feedback components required for effective feedback are needed. This is likely to involve a series of studies designed to isolate one feedback component at a time, with all other key influences on performance standardised, in order to identify, replicate and validate the conditions that are most effective in helping learners to improve, across different contexts.

Summary

We systematically collated the available evidence regarding the impact of verbal face-to-face feedback on health professionals' performance, compared with no or alternative feedback. In a meta-analysis we found that verbal face-to-face feedback substantially enhanced workplace task performance compared to no feedback SMD 0.7 (95% CI: 0.37-1.03; P<0.001).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, incorporating multiple distinct components. More robust evidence is required and future research should focus on systematically analysing components that maximise the effects of feedback.

Declarations of interest: The authors report no declarations of interest.

Author Contributions: CJ and JK designed the review, CJ created the search strategy and undertook the literature searches, CJ and MW undertook study screening, data extraction and quality assessment with oversight from JK, CJ and JK undertook data analysis and interpreted the findings, CJ prepared the manuscript and all authors contributed to the final version.

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

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: The authors report no competing interests.

Patient consent: Not required.

Ethics Approval: Not required.

Data sharing statement: Data access can be requested by contacting the corresponding author.

Figure and Table Legends

Figure 1: PRISMA flow diagram.

Figure 2: Risk of bias graph

Figure 3: Risk of bias summary

Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback on performance, compared to no feedback.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback on performance, compared to no feedback.

Figure 6: Summary of findings table for the effect of verbal face-to-face feedback on performance, compared to no feedback.

Figure 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A) compared to alternative feedback (Feedback B), on performance.

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

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

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(5):645-9. doi: 10.1097/acm.0000000000001048 [published Online First: 2015/12/18]
3. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34(3):205-14. doi: 10.3109/0142159x.2012.652239 [published Online First: 2012/03/01]
4. Holmboe ES. Realizing the promise of competency-based medical education. *Academic Medicine* 2015;90(4):411-3. doi: 10.1097/acm.0000000000000515 [published Online First: 2014/10/09]
5. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association* 2009;302(12):1316-26. doi: https://dx.doi.org/10.1001/jama.2009.1365
6. Archer JC. State of the science in health professional education: effective feedback. *Medical Education* 2010;44:101-08. doi: 10.1111/j.1365-2923.2009.03546.x

7. Bing-You RG, Hayes V, Varaklis K, et al. Feedback for learners in medical education: what is known? A scoping review. *Academic Medicine* 2017;92(9):1346-54. doi: https://dx.doi.org/10.1097/ACM.0000000000001578
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 Medical Education* 2016;16(1):96. doi: 10.1186/s12909-016-0613-5
9. Watling C, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ* 2019;53(1):76-85. doi: 10.1111/medu.13645 [published Online First: 2018/08/04]
10. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist* 2002;57(9):705-17. [published Online First: 2002/09/20]
11. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine* 2015;90(5):609-14. doi: https://dx.doi.org/10.1097/ACM.0000000000000560
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. *Medical Teacher* 2011;33(12):961-73. doi: https://dx.doi.org/10.3109/0142159X.2011.595435
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 Medical Education* 2019;19(1):129. doi: 10.1186/s12909-019-1524-z
14. Hattie J, Timperley H. The Power of Feedback. *Review of Educational Research* 2007;77(1):81-112. doi: 10.3102/003465430298487
15. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
16. Molloy E, Boud D. Changing conceptions of feedback. In: Boud D. ME, ed. Feedback in higher and professional education London: Routledge 2013:11-33.
17. Ende J. Feedback in clinical medical education. *Journal of the American Medical Association* 1983;250(6):777-81. doi: 10.1001/jama.1983.03340060055026
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. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
19. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
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(4):264-9, w64. doi: 10.7326/0003-4819-151-4-200908180-00135 [published Online First: 2009/07/23]
21. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration 2011 [Available from: www.handbook.cochrane.org.
22. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
23. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List. *Journal of Clinical Epidemiology* 1998;51(12):1235-41. doi: 10.1016/S0895-4356(98)00131-0
24. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal* 2011;342:d549.
25. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b

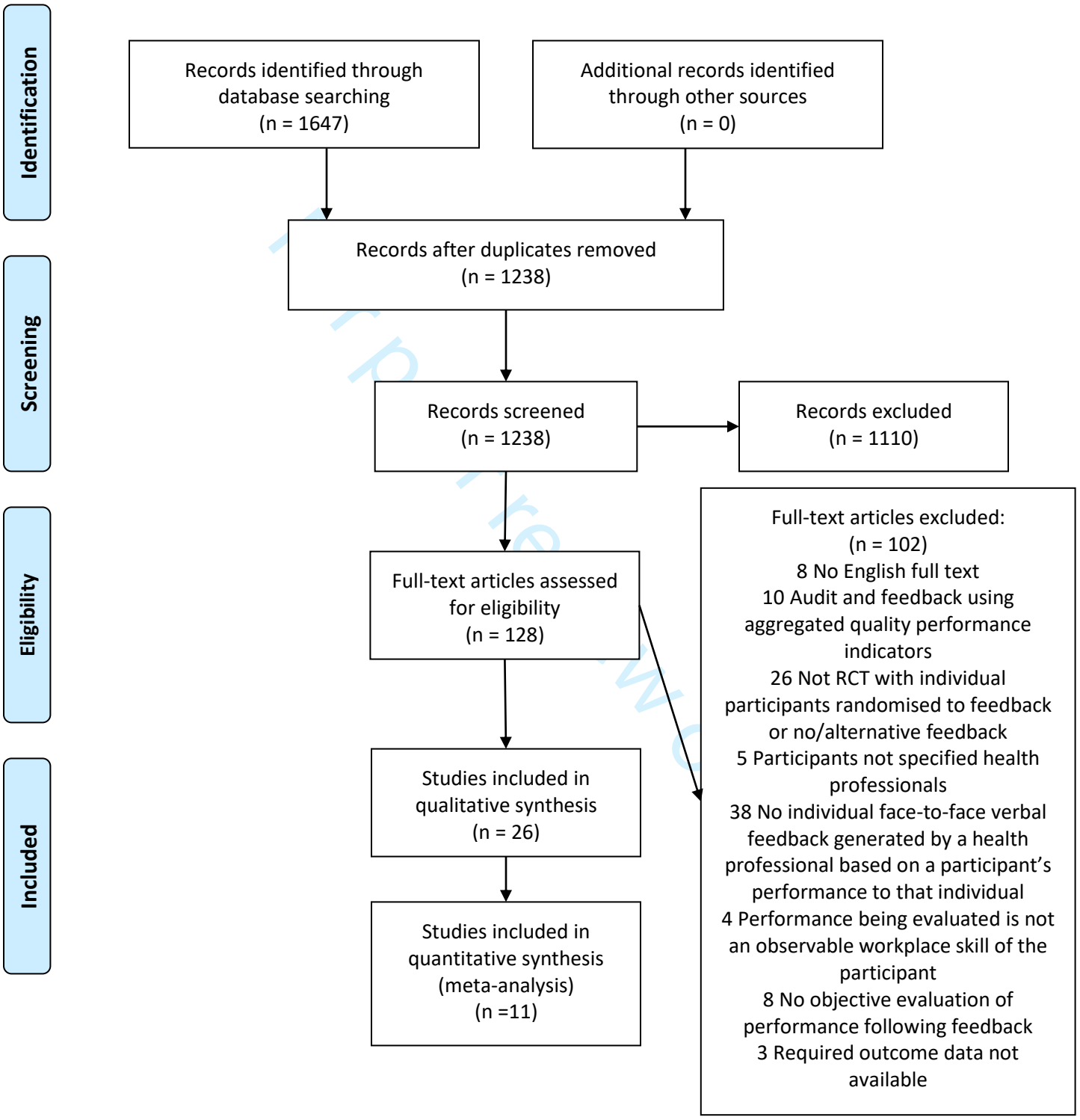
26. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
27. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
28. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
29. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
30. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
31. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
32. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
33. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
34. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
35. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
36. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
37. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
38. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
39. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5
40. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine* 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44
41. 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(1):322-28.
42. 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 Medical Simulation 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165

43. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
44. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e
45. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
46. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
47. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009
48. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397
49. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209
50. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
51. Sullivan GM, Feinn R. Using Effect Size -or Why the P Value Is Not Enough. *Journal of Graduate Medical Education* 2012;4(3):279-82. doi: 10.4300/JGME-D-12-00156.1
52. Coe R. It's the effect size, stupid: What effect size is and why it is important. A paper presented at the Annual Conference of the British Educational Research Association 2002. 2002
53. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
54. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
55. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
56. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
57. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.
58. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
59. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360
60. Elliott SN. Educational psychology: effective teaching, effective learning. 3rd ed. Boston: McGraw-Hill 2000.

61. Wadsworth BJ. Piaget's theory of cognitive and affective development: foundations of constructivism., 5th edn. White Plains: Longman Publishing 1996.
62. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
63. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
64. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.
65. 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 ed. Oxford: Wiley Blackwell 2014:7-29.
66. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227-68. doi: 10.1207/s15327965pli1104_01
67. Belfield C, Thomas H, Bullock A, et al. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;23(2):164-70. doi: 10.1080/01421500020031084 [published Online First: 2001/05/24]

Figure 1: PRISMA flow diagram for systematic review of verbal face-to-face feedback compared to no or alternative feedback.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

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

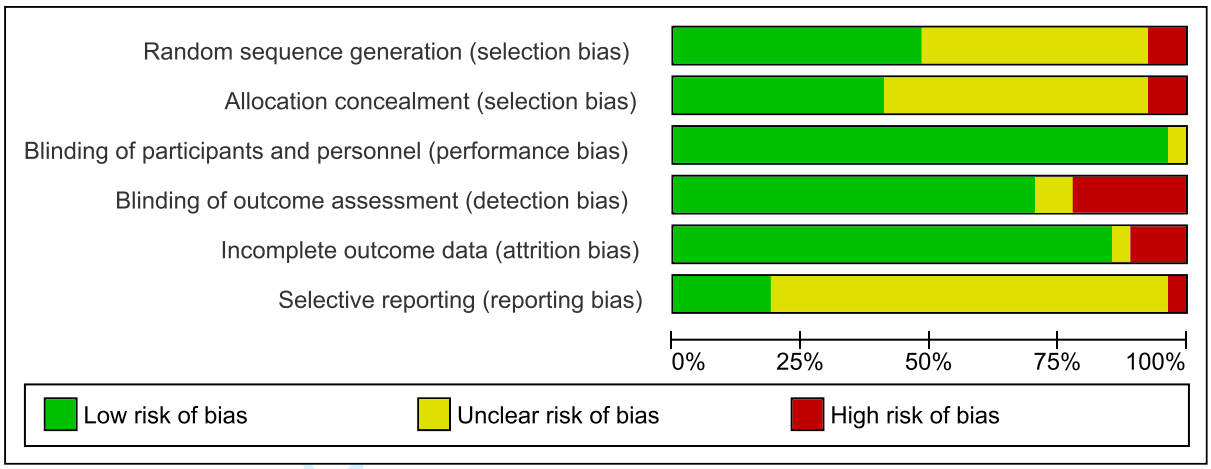


Figure 3: Risk of bias summary: review authors' judgements about each risk of bias item for each included study

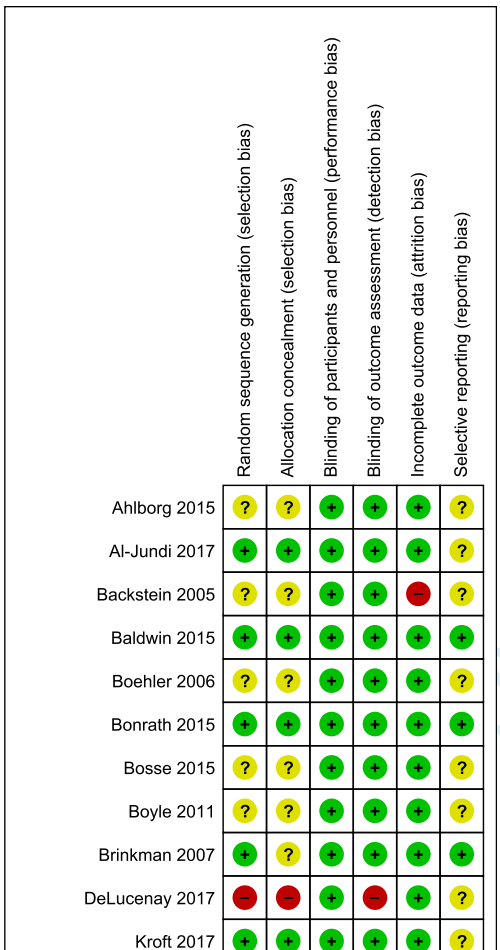
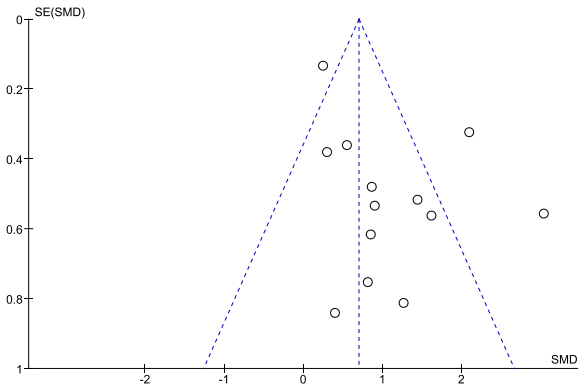
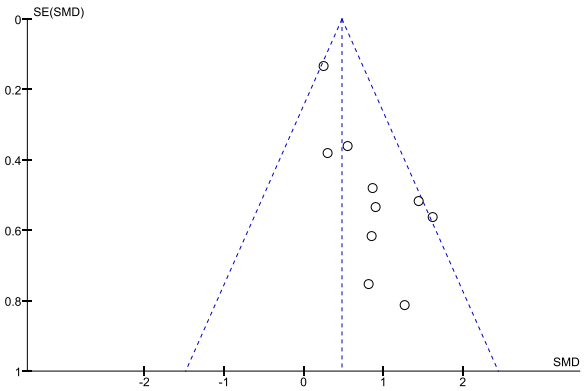


Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback on performance, compared to no feedback.

a) all included studies



b) excluding studies at high risk of bias



Abbreviations:
SE = standard error; SMD = standardised mean difference

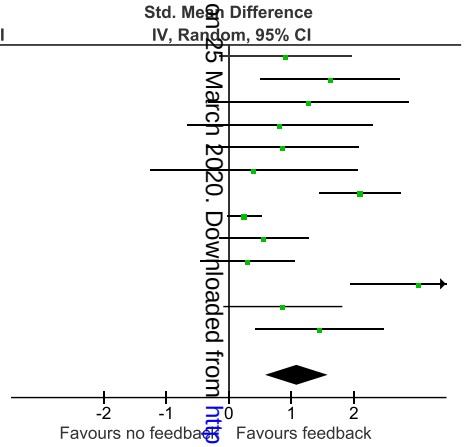
Footnote:
Meta-analysis calcuated using a fixed effects model.
The dotted vertical line represents the overall effect estimate and the dotted slanted lines represent the 95% confidence interval lines.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback on performance compared to no feedback

a) All included studies

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.6%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.3%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	5.3%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	5.7%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.9%	0.85 [-0.35, 2.06]
O'Connor 2008	-1.96	1.25	3	-2.5	0.9	3	5.1%	0.40 [-1.25, 2.04]
Olms 2016	-1.1	1.8	30	-7.4	3.8	30	9.6%	2.09 [1.45, 2.73]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	11.0%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	9.2%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	9.1%	0.30 [-0.44, 1.05]
Vafei 2017	17.73	1.09	15	12.13	2.29	15	7.4%	3.04 [1.95, 4.13]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	8.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	7.7%	1.44 [0.43, 2.46]
Total (95% CI)			**253**			**235**	**100.0%**	**1.09 [0.59, 1.59]**

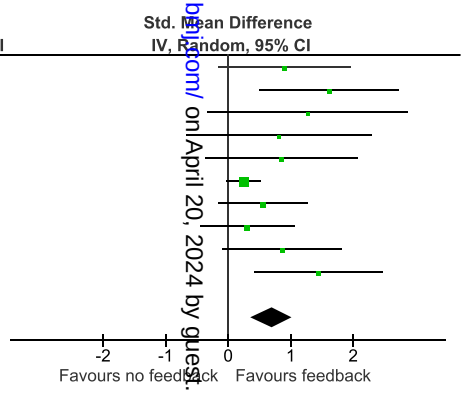
Heterogeneity: Tau² = 0.58; Chi² = 54.26, df = 12 (P < 0.00001); I² = 78%
Test for overall effect: Z = 4.26 (P < 0.0001)



b) Excluding studies at high risk of bias (sensitivity analysis)

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.7%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.1%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	3.8%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	4.4%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.2%	0.85 [-0.35, 2.06]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	27.9%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	13.3%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	12.4%	0.30 [-0.44, 1.05]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	9.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	8.1%	1.44 [0.43, 2.46]
Total (95% CI)			**205**			**187**	**100.0%**	**0.70 [0.37, 1.03]**

Heterogeneity: Tau² = 0.08; Chi² = 13.64, df = 9 (P = 0.14); I² = 34%
Test for overall effect: Z = 4.14 (P < 0.0001)



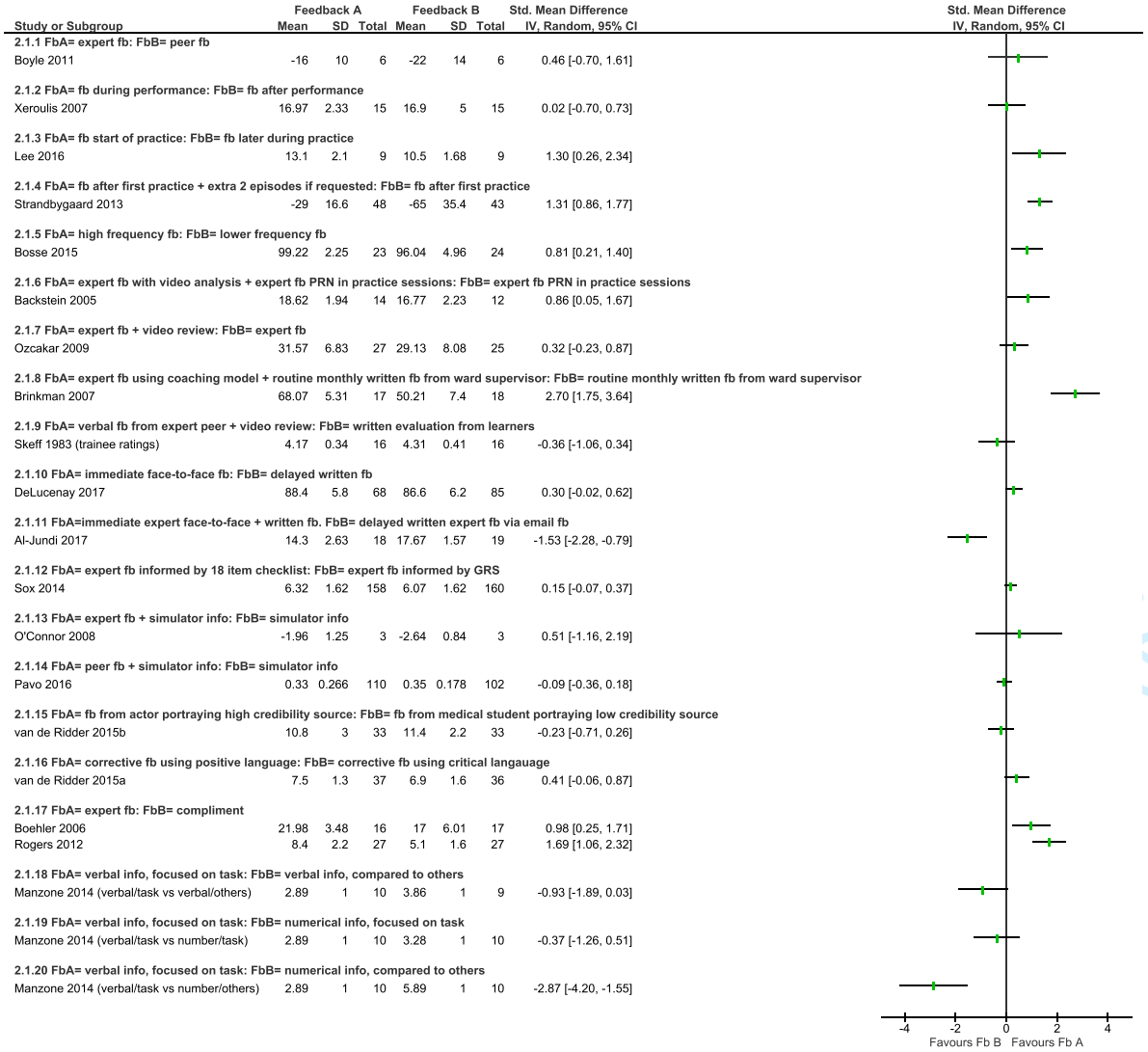
Abbreviations:
SD = standard deviation; CI = confidence interval; SMD = standardised mean difference

Footnotes:
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

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

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 (0.37-1.03) higher than mean scores for the control group	Number of participants 392 (8 studies)	⊕⊕⊕O[a,b,c] Moderate	A SMD of 0.7 indicates a substantial improvement in performance
CI = Confidence interval; SMD= standardised mean difference				
[a]High risk of bias due to lack of allocation concealment and selective reporting of outcomes. [b]Unexplained heterogeneity [c]High probability of publication bias				
GRADE Working Group grades of evidence ⊕⊕⊕⊕ High: We are very confident that the true effect lies close to that of the estimate of the effect. ⊕⊕⊕O Moderate: We are moderately confident in the effect estimate, the true effect is likely to be close to this but there is a possibility that it is substantially different ⊕⊕OO Low: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect ⊕OOO Very Low: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.				

Figure 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance



Abbreviations:
Fb= feedback; GRS= global rating scale; info= information; PRN= 'as required'; SD = standard deviation; CI=confidence interval

Footnotes:
[1]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

Database(s): Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:

#	Searches	Results
1	*Feedback/	6031
2	Feedback, Psychological/	3311
3	Formative Feedback/	467
4	(feedback adj3 (effective or formative or constructive or quality or clinical or performance)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4860
5	1 or 2 or 3 or 4	13233
6	exp Health Personnel/	470058
7	exp Health Occupations/	1648689
8	exp Dentistry/	386159
9	exp Social Work/	17331
10	exp Psychology/	66579
11	Occupational Therapy/	13213
12	Radiotherapy/	42757
13	Radiography/	334082
14	Mentors/	9949
15	exp Students, Health Occupations/	60760
16	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2722708
17	clinician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	171551
18	(health* adj2 (staff or personnel or faculty or provider* or worker* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	303659
19	doctor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104240
20	physician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	495037
21	(medical adj3 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	133207
22	general practitioner*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45015
23	(general pract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45488
24	(family adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21549
25	(primary care adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10661
26	(primary health* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2156
27	(registrar or registrars or senior house officer* or resident or residents or hospital medical officer* or intern or interns or house officer*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	129452
28	dentist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	115221
29	(dent* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21015
30	nurs*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	629304
31	(midwife or midwives).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	18155

32	(midwife* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	965
33	(allied health adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	12936
34	physiotherapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5348
35	physical therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5123
36	(physiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	731
37	(physical therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	647
38	occupational therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4500
39	(occupational therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1058
40	speech therap*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7630
41	(speech therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	22
42	speech language therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	136
43	(speech language therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	14
44	speech pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	656
45	(speech pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	31
46	speech language pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1561
47	(speech language pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	105
48	dietician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1352
49	(dietetic* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	603
50	podiatrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	626
51	(podiatry adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	19
52	chiropodist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	119
53	(chiropody adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2
54	social worker*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8079
55	(social work* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8517

56	psychologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	11986
57	(psychology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1532
58	(osteopath* or osteopathic physician*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	6480
59	(osteopath* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	296
60	chiropractor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1148
61	(chiropract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	241
62	pharmacist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	27316
63	(pharmac* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7184
64	optometrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1764
65	(optometr* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	290
66	(Radiographer* or radiological technologist* or radiation therapist* or radiotherapist* or radiation therapy technologist*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2449
67	(radiograph* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	267
68	(radiation therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	53
69	(radiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	100
70	(supervisor* or tutor* or trainer* or educator* or teacher* or mentor* or preceptor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104110
71	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70	1846616
72	16 or 71	3625236
73	exp Education, Professional/	284462
74	exp Educational Measurement/	137138
75	exp Professional Practice/	247602
76	exp Simulation Training/	6239
77	(effect* or evaluat* or outcome* or assess* or measur*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10754772
78	73 or 74 or 75 or 76 or 77	11134144
79	randomized controlled trial.pt.	505181
80	controlled clinical trial.pt.	100406
81	randomized.ab.	391590
82	randomly.ab.	266043
83	systematic review.ab,ti.	85419
84	79 or 80 or 81 or 82 or 83	966230
85	5 and 72 and 78 and 84	821
86	limit 85 to (english language and humans)	809

Supplementary online material. Appendix 2

COMPARISON 1: THE EFFECT OF VERBAL FACE-TO-FACE FEEDBACK ON PERFORMANCE, COMPARED TO NO FEEDBACK.

Results

Included studies

Participants

Participants included 290 (60%) medical students in four studies,[1-4] 60 (12%) dental students in one study[5] and 138 (28%) doctors (doctors training in surgery in three studies,[6-8] training in obstetrics and gynaecology in one study[9] and training in emergency medicine in one study,[10] and physicians in one study[11]).

Participants were novices to the assessed task in five studies (5/11, 45%);[1, 2, 4, 5, 7] and had prior experience in six studies.[3, 6, 8-11]

Workplace tasks and Settings

All studies evaluated performance of a discrete task; there were no longitudinal evaluations. The task occurred in simulation settings in seven studies (7/11, 64%) and clinical practice in four studies (4/11, 36%). The task was a surgical procedure in seven studies (7/11, 64%). Five studies involved simulated surgical tasks including bench top models for knot tying[4] and forming a bowel anastomosis;[8] using a laparoscopic simulator for suturing and knot tying;[2] and using a virtual reality (VR) simulator for laparoscopic surgery[1] and endovascular surgery.[7] Two studies involved laparoscopic surgery in clinical practice.[6, 9] The remaining four studies evaluated simulated matching of tooth colour in a dental school,[5] simulated cardiopulmonary resuscitation (CPR),[3] chest ultrasound for emergency trauma patients[10] and teaching skills in clinical practice.[11]

Feedback Interventions

The feedback source involved a subject expert in all comparisons except two, including one that compared peer feedback with no feedback,[3] and one that compared expert feedback, peer feedback and no feedback.[7] Feedback occurred while the participant performed the task (during) in one study,[3] both during and directly afterwards in two studies,[1, 2], directly afterwards in four studies,[5, 7, 9, 10] after a delay in three studies[6, 8, 11] and one study compared feedback during, feedback directly afterwards and no feedback.[4] In addition to evaluative performance information (as per inclusion criteria), the feedback included corrective advice in all studies except one[10] and one where it was unclear.[7] Feedback included additional information from a simulator in three studies,[1, 2, 7] a video of the participant's performance in two studies[6, 11] and written performance information in two studies.[5, 11]

Teaching and Practice

In addition, instruction and expert demonstration of the task was provided in six studies (6/11, 55%), including all five studies involving novice participants[1, 2, 4, 5, 7] and one study that involved CPR for medical students, many of whom had previously attended a course.[3] The other five studies involved doctors working in clinical practice; in these studies, no instruction or expert demonstration was included within the research intervention but may or may not have occurred during the course of routine work during that time. One study involved physicians' teaching on ward rounds[11] and the other four studies assessed tasks by doctors training in relevant specialties.[6, 8-10]

The amount of practice varied substantially between different studies, for both simple and complex tasks. For example, comparing two studies that involved simple surgical knot tying: in Xeroulis,[4] participants practiced 18 times in one session and in O'Connor,[2] they could practice up to an hour a day, for 24 days. Looking at more complex surgical procedures, such as simulated surgery using a virtual reality (VR) simulator: in Ahlborg,[1] participants had two

practice attempts at the simulated surgery (laparoscopic salpingectomy) and in Boyle,[7] participants had five attempts at the simulated surgery (renal artery angioplasty and stenting) before the performance evaluation.

Intervention period

The intervention period ranged from one day (most common) up to two months.[6] Nine (9/11, 82%) studies involved a single session (involving one episode of feedback in five studies[5, 8-11] and multiple episodes of feedback in four studies[1, 3, 4, 7]). Two studies (2/11, 18%) involved multiple feedback sessions over a longer period: one study[6] included approximately four coaching sessions regarding bariatric surgery across a two month surgical attachment, and another[2] included almost daily one hour practice sessions for laparoscopic suturing, with feedback throughout each one, over four weeks.

The timing of the post-feedback performance assessment, in relation to the intervention, differed. It occurred directly following the intervention in seven studies: at the end of the single session in five studies[1, 3, 4, 7, 9] and at the end of an extended intervention period in two studies.[2, 6] In the other four studies, the post-feedback performance assessment occurred some weeks after the intervention was completed but while relevant exposure to possible teaching and/or practice opportunities continued. Olms[5] included a single feedback session, with the final evaluation two weeks later, in the midst of a routine one month university teaching unit on tooth shade matching. Skeff[11] arranged a single coaching session on ward round teaching in the middle of physicians' four week ward duty, with the final evaluation post-performance evaluation at the end. Soucisse[8] also organised a single coaching session for surgical residents, with the final evaluation occurring three weeks later. Vafaei[10] involved a single workplace-based assessment with feedback for doctors training in emergency medicine on chest ultrasound for emergency trauma patients, followed by a two month period of routine clinical work before the post-feedback assessment.

Research funding

Regarding research funding, one study[3] that focused on cardiopulmonary resuscitation (CPR) quality, was loaned a device (used to measure CPR parameters and provide automated feedback to participants) for the period of the study by Philips but the company was not otherwise involved in the research; five studies received funding from independent institutions,[1, 4, 6, 9, 11] three studies did not receive any funding[5, 7, 10] and two studies did not report information on funding.[2, 8]

Risk of bias

Five trials described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'. [3, 5, 6, 8, 9] The other six trials simply stated participants were 'randomised' and had no information on allocation concealment, so we rated these studies as 'unclear'. We analysed baseline performance because, although randomisation removes the need to check comparability in baseline task performance for intervention and comparison groups, it may be useful to check this when participant numbers are small and performance improvement is more likely when baseline performance is low.[12] Seven studies reported no statistically significant differences between baseline performances for the comparison groups.[4, 5, 8-11] and four studies did not report baseline task performance.[1-3, 7] The participants and research team members were not blinded in any included studies because the intervention involved feedback between a research team member and a participant, consistent with most education interventions. However, in all included studies, we thought this was not likely to influence the outcome (post-intervention performance assessment) because implementation and adherence to the intervention were not affected. In eight studies the outcome was assessed by either blinded assessors who rated videos of the participants' performance[4, 6-9, 11] or by a machine (simulator or CPR machine),[1, 3] so we rated these as 'low risk' of bias. In three studies, the feedback provider and outcome assessor appeared to be the

same person, so these were rated as 'high risk'.[2, 5, 10] Across all the studies, the follow up rate for each group was at least 85%. Only two studies had a prior published protocol in addition to reporting all outcomes as planned.[6, 8] For all other studies, it could not be ascertained if outcomes had been selectively reported, so these were rated as 'unclear', except one. This one study was rated as 'high risk' for selective outcome reporting because it did not include the expected information on performance post-intervention.[2]

In summarising the risk of bias across domains within each study, two studies had all domains rated 'low risk, so these were rated low risk.[6, 8] Six studies had at least one domain with 'unclear' risk but no 'high risk' ratings, so these were rated as 'unclear' risk of bias.[1, 3, 4, 7, 9, 11] Three studies had at least one domain at high risk of bias, so we judged these studies to be at 'high risk' of bias.[2, 5, 10]

Certainty of evidence

For the comparison of verbal face-to-face feedback compared to no feedback, excluding studies at high risk of bias, we graded the quality of evidence for the outcome of 'objective assessment of a health professional's performance'. The risk of bias was rated as 'unclear' across multiple included studies and the overall body of evidence indicated this was likely to seriously alter the results, so we downgraded the overall evidence by one level. The two aspects that were most influential on our decision were the lack of allocation concealment and prior published protocols to ensure selective reporting of outcomes did not occur. Participant and research team member blinding was not possible due to the intervention. However, this had limited impact on the selected outcome 'objective assessment of performance', as no changes occurred in intervention implementation or adherence as a consequence of this lack of blinding.[13] We judged the results to be directly applicable to our review question and therefore the evidence was not downgraded for indirectness. We judged there to be some variability across studies and therefore downgraded the evidence due to inconsistency by one level. This was based on the methodological and statistical heterogeneity which was not explained by subgroup analysis. We judged the effect size to be sufficiently precise and therefore did not downgrade the evidence for

imprecision of results. This was based on sufficient numbers of participants (392 when studies with high risk of bias were excluded) and a consistent beneficial effect, indicated by the confidence interval for the overall effect estimate not crossing zero and all individual studies showing a beneficial effect with substantial overlap in their confidence intervals. Finally, we judged that there was likely to be a systematic overestimation of the underlying beneficial effect of feedback because we strongly suspected publication bias (see Funnel plot 5b) and therefore we downgraded the evidence by one grade.[14]

In summary, combining all five GRADE criteria for assessing the certainty of evidence, we downgraded the overall rating by one, from high to moderate. We judged that the quality of the evidence contributing to the effect estimate of 0.70 in the comparison of verbal face-to-face feedback to no feedback after excluding studies with a high risk of bias, was moderate. Hence, we are moderately confident in the effect estimate; the true effect is likely to be close to this but there is a possibility that it is substantially different.

COMPARISON 2: THE EFFECT OF VERBAL FACE-TO-FACE FEEDBACK ON PERFORMANCE, COMPARED TO ALTERNATIVE FEEDBACK.

Results

Included studies

Participants

Included studies involved medical students (1076, 55%) in 14 studies,[2-4, 15-25] mixed health professional students (640, 32%) in one study,[26] pharmacy students (153, 8%) in one study[27] and doctors (105, 5%) in four studies.[7, 11, 28, 29]

Participants were novice to the task in 11 studies (11/20, 55%). Three studies documented prior experience: one study involved attending physicians teaching on ward rounds with a range of experience[11] and two studies documented previous training including CPR[3] and history taking

and communication skills in medical students.[24] The remaining six studies did not report this information. One of these studies[26] involved teaching CPR to first year health professional students across a mix of disciplines, some of whom may have had prior experience. One study[29] involved evaluating professional and clinical skills in first year paediatric residents who likely had relevant training as medical students. In two of these studies, the baseline performance of junior medical students' attempting surgical knot tying was poor, which suggest limited prior experience.[16, 17] In the last two studies there was no information on prior experience: one assessed a simulated medication consultation by third year pharmacy students[27] and another[23] assessed case presentation skills in third year medical students in their paediatric attachment.

Workplace tasks and Settings

All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[11, 29] Three studies evaluated performance in a clinical practice setting (involving teaching skills[11] professional and communication skills[29] and oral case presentations[23]) and the remaining 17 assessed performance in a simulated environment.[2-4, 7, 15-22, 24-28] Simulated surgical tasks included suturing and/or knot tying,[2, 4, 15-17] bench top surgical procedures such as vascular anastomosis,[28] flexible ureteroscopy for urolithiasis,[19] renal artery angioplasty and stent,[7] or surgery using a VR simulator for a laparoscopic salpingectomy.[25] Simulated critical care tasks included basic life support(BLS)/CPR[3, 26], intubation[20] and pharmacist-patient consultation.[27] The remaining simulated tasks included a hearing test,[21, 22] simulated patient consultation[24] and nasogastric tube insertion.[18]

Interventions

Each study included at least one verbal face-to-face feedback group, in accordance with the inclusion criteria.

Some studies investigated straightforward variations in feedback, including differences in frequency (low or high[18]), stage of practice (early or late[19]), different feedback models ('learning conversation' compared with 'feedback sandwich' frameworks[26]), source expertise (expert or peer[7]) and expert feedback compared to compliments.[16, 17] Another collection of studies explored the effect of adding expert feedback to other interventions, such as in addition to simulator performance data[3, 25] or to written feedback;[11, 29] or adding expert review of a participant's performance video to a practice session in which expert feedback was available.[28] One study[15] compared verbal feedback by an expert who had just directly observed the performance, with written feedback emailed later that day by another expert who watched a video of the performance. Other studies explored more complex phenomena. One study[20] compared two feedback variations in different combinations across four groups. One variation compared an evaluative verbal comment from an expert, to a written numerical performance rating. The second variation involved an individual comparing their performance evaluation to either their own previous attempts (highlighting individual progress) or to expected performance at student, resident or specialist level (comparison with others). Another research group investigated two complex influences in separate studies. One study[22] examined how the credibility of the feedback provider (high or low credibility) influenced learner outcomes. The other study[21] examined the effect of phrasing corrective information in different ways, so in one intervention corrective information was framed within a positive phrase whereas in the other, it was framed within a critical phrase.

Research funding

One study[3] was loaned a device by Philips as detailed earlier, seven studies received funding from independent institutions,[4, 11, 18, 20, 23, 25, 29] six studies did not receive any funding[7, 16, 19, 21, 22, 24] and six studies did not report information on funding.[2, 15, 17, 26-28]

Risk of bias

The risk of bias assessment for the comparison of verbal face-to-face feedback to alternative feedback is presented in Figure 3. Seven described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[3, 15, 21-23, 25, 26] Two studies adequate random sequence generation, which we rated 'low risk' but had insufficient information on allocation concealment, which we rated 'unclear risk'.[24, 29] The remaining studies simply stated participants were 'randomised' and had insufficient information on allocation concealment, both of which we rated 'unclear risk'. Two studies described inconsistencies with randomisation, so these were rated 'high risk' of bias for sequence generation and allocation concealment.[20, 27] There was unequal baseline performance between groups reported in one study[22] and identified from another study's data (obtained from authors)[15]. No statistically significant differences in baseline performance between groups were reported in seven studies [4, 11, 16, 18, 19, 22, 29] and baseline performance was not reported in eleven studies.[2, 3, 7, 17, 20, 23-28] None could blind participants or research team members due to the face-to-face feedback interventions. However we thought this was not likely to influence the outcome as implementation and adherence to the intervention were not affected in all studies, which were rated 'low risk', except one in which some participants may not have experienced the intervention they were allocated to, so it was rated 'unclear'.[23] The outcome was assessed by blinded assessors or machines in all studies, which were 'rated low risk' except two studies that did not explicitly describe blinded assessors, which were rated 'unclear'[11, 22] and four studies that seemed to have assessors who were aware of participant allocation, so these were rated 'high risk'.[2, 24, 27] All had high proportions of participant completion data except three[21, 22, 28] and one report provided insufficient information[17]. Three studies had prior published protocols and reported all outcomes as planned, so they were rated 'low risk'.[25, 26, 29]

All of the others did not have a prior published protocol but did present outcomes as expected and were rated as 'unclear'.[3, 4, 7, 11, 15-24, 27, 28] except one study which was rated as 'high risk'.[2]

In summarising the risk of bias across domains within each study, two studies were rated as low risk[25, 26] as all domains were rated as 'low risk' of bias, seven studies were rated as 'high risk' because at least one domain was rated as 'high risk'[2, 20-22, 27, 28] and the remaining studies were rated as 'unclear' as they had at least one domain with 'unclear' risk but no 'high risk' ratings.

References

1. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
2. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
3. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
4. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
5. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
6. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214

7. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
8. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
9. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
10. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
11. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
12. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
13. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
14. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration 2011 [Available from: www.handbook.cochrane.org.
15. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
16. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
17. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009
18. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5
19. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
20. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397
21. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
22. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
23. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209

24. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
25. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e
26. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
27. 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(1):322-28.
28. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
29. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine* 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	P1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	P2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	P4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	P5-6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	P6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	P6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	P6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	P6-7 & Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	P6-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	P6-8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	P7, Tables 1 & 2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	P7-8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	P9

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	P9-10

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	P7-8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	P9-10
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	P10 & Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	P10-12 & 17-20 Tables 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	P14-16, & 20 & Figures 2 & 5
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures 2 & 5
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	P9 & Figures 2-4
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	P9-11, 20 Figures 2 & 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	P9-10, Figures 2,4 & 5
DISCUSSION			

PRISMA 2009 Checklist

Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	P21-22
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	P26
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	P26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	P27

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030672.R2
Article Type:	Original research
Date Submitted by the Author:	05-Jan-2020
Complete List of Authors:	Johnson, Christina; Monash Health, Monash Doctors Education ; University of Melbourne, Department of Medical Education, Melbourne Medical School Weerasuria, Mihiri; Monash Health Keating, Jennifer; Monash University, Department of Physiotherapy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	MEDICAL EDUCATION & TRAINING, feedback, health professions education, formative feedback

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Short title: Systematic review and meta-analysis on feedback

Christina E. Johnson MBChB, MHPE, FRACP

ORCID ID: 0000-0002-4209-8419

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, Australia.

Mihiri P. Weerasuria MBBS (Hons)

ORCID ID: 0000-0001-6897-3982.

Monash Health, Melbourne, Australia.

Jennifer L. Keating PhD, GradDipManip PT, BAppSci (PT)

ORCID ID: 0000-0003-3161-4964

Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia.

Correspondence should be addressed to Christina Johnson, Monash Doctors Education, McCulloch House, Monash Health, Clayton Road, Clayton, Victoria 3168, Australia. Telephone: 613-9594 3743. Email: 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 1st 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 to no feedback or alternative feedback, and available as full text publications in English. The certainty of evidence was assessed using the GRADE approach. For feedback compared to 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 to 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 0.37-1.03; $P<0.001$) in favour of feedback. The calculated SMD prediction interval was -0.06 to 1.46. For feedback compared to alternative feedback, studies could not be pooled due to substantial intervention

and design 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: We are moderately confident that verbal face-to-face feedback in the health professions substantially enhances workplace task performance, compared to no feedback. However further research is needed, as we found a lack of high-quality trials that clearly reported key components likely to influence performance. To build a robust evidence-base, and identify best practice in feedback, a standardised approach to investigations and reporting is required.

PROSPERO 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 to 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.

• For the meta-analysis, not all studies that met the inclusion criteria clearly reported the data required for pooling, so for some studies data were obtained from the author or estimated by calculating these using available data or reading off graphs; otherwise, the study was excluded.

Keywords: feedback, effective feedback, formative feedback, systematic review, meta-analysis, health professions education

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 (for example, from a simulator or within an online learning module). The unique potential benefit of face-to-face verbal feedback is the opportunity for i) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to co-construct insights and strategies for improvement and ii) inter-personal 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 said, "Without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all."[17] 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 which investigated the impact of audit and feedback. In 2006, Veloski et al 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.[18] 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 commonest data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers et al updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared to usual care.[19] 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% (interquartile range 0.5-16%), the 75th centile of 16% suggested 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 to 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 e.g. 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) Registration ID CRD42017081796.

Eligibility criteria for considering studies for this review

We included randomised controlled trials in which individual health professionals were randomised to feedback, compared to 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 pre-determined 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 i) verbal face-to-face feedback compared with no feedback and ii) verbal face-to-face feedback compared to alternative feedback.

Performance following feedback interventions had to be objectively assessed. To isolate the effects of feedback, 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 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 1st February 2019. We also searched the reference lists of systematic reviews and included studies.

Data collection and analysis

Selection of studies

One review author (CJ or MW) screened titles to exclude clearly irrelevant reports. Two authors (CJ and MW) independently screened remaining abstracts to identify potential eligible studies, then independently assessed the full text. Decisions were compared using Covidence (on-line 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 (JK).

Data extraction and management

One review author (CJ) used a pre-piloted standardised form to extract data from included studies and another author (MW or JK) 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.

Assessment of risk of bias in included studies

The risk of bias was independently assessed by two authors (CJ and JK) 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 (for example means (standard deviations SD) or medians (interquartile range IQR)). The effect was quantified using the standardised mean difference 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).[21]

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 standardised mean difference (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.[24]

We rated the overall certainty of evidence for the outcome using the GRADE approach (Chapter 12, Cochrane Handbook),[21] which considers within-study risk of bias, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias.[21] 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.

[Figure 1 here]

Comparison 1: The effect of verbal face-to-face feedback, compared to 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 to no feedback on the objective evaluation of a workplace task. Seven (64%) reports were published in the last five years since 2014. The studies were conducted in Europe (4),[25-28] Canada (4),[29-32] the USA (2)[33, 34] and Asia (1).[35]

There were 488 participants, including 196/366 (53.6%) males from seven studies that reported gender data.[25, 27-30, 33, 35] Participants included 290 (60%) medical students in four studies,[25, 28, 32, 33] 60 (12%) dental students in one study[27] and 138 (28%) doctors in four studies.[26, 29-31, 34, 35] 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 [29, 30, 34, 35] and a simulation environment in seven studies (7/11, 64%).[25-28, 31-33] 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 - Appendix 2).

Table 1: Summary of available data on characteristics of trials included in the comparison of verbal face-to-face feedback (intervention) compared to 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** Additional information	Source		Timing			Content					**Study outcomes**[a] SMD (95% CI) Bolded text indicates significant effects ($p<0.05$) All in favour of feedback
					Subject Expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info[b]	Performance video	Written performance info	
Ahlborg 2015 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 x practice trials. Performance evaluation: end of session.	2 x fb episodes. Fb given by expert i) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus ii) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓			0.91 (-0.14 – 1.95)
Bonrath 2015 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.	4 (approx.) x 25 min fb episodes. Fb given by expert using specific coaching model[b] + video review of learner operating + video exemplars of good/poor technique. Effectiveness of strategies	✓				✓	✓	✓		✓		**1.62** (0.52 – 2.72)

			Performance evaluation: end of clinical attachment.	reviewed at subsequent session.										
Boyle 2011 (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 x practice trials Performance evaluation: end of session	5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' + simulator output information.	✓			✓		?	?	✓			1.27 (-0.32 – 2.87)
Boyle 2011 (peer fb)	Same as above	Same as above	Same as above	5 x fb episodes. Peer discussed simulator output, any task errors & teaching instructions given at start.		✓		✓		✓	✓	✓			0.81 (-0.66 – 2.29)
Kroft 2017 Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy)	Doctors training in O&G PGY2-6 33% M	Intervention duration: 1 x 15min practice using laparoscopic salpingectomy module on VR surgical simulator within 1h of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1 x fb episode from expert directly after VR simulator practice. Fb 'standardized and given in an evidence based fashion to optimise effectiveness' & included '3 constructive recommendations based on performance'.	✓			✓		✓	✓				0.85 (-0.35 – 2.06)
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	Medical students UGY1-2 44% M	Intervention duration: 4 wk. 2h instruction + practice suturing & knot tying until able to do it easily. Then instruction on laparoscopic surgery + expert	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 – 2.04)

			demonstration video of task tying, followed by 30 mins familiarisation with equipment. Practice: 1h daily, 6 days per week for 4 weeks. Performance evaluation: combined assessment of each attempt throughout intervention.											
Olms 2016 Germany	Simulated colour matching teeth	Dental students UGY3	Intervention duration: 1 session Study conducted during 10 wk routine university module on matching tooth shades involving variety of teaching + practice opportunities. Performance evaluation: 2 wks after intervention (within one university module).	1 x 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 – 2.73)
Pavo 2016 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 x 2h additional training session: instructional video + training using	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 & depth, hand position,		✓	✓			✓	✓				0.25 (-0.02 – 0.51)

			modified Peyton 4 step approach.[d] Brief practice (few mins) in pairs using a manikin. Performance evaluation: end of session.	decompression & hands-off time. Instructional video for intervention group had demonstrated this.										
Skeff 1983 USA	Clinical teaching skills during ward round	Physicians	Intervention duration:1 session in the middle of 4wk ward duty. At mid & 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 wk later, at end of ward duty	1 x 60 min fb discussion with peer, including video review, trainee ratings & self-assessment to enable physician to identify strengths & devise solutions to problems.	✓				✓	✓	✓		✓	✓	0.56 (-0.15 – 1.27)
Soucisse 2017 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 wk later	1 x 30min expert fb sometime after baseline performance with video review of baseline performance + coaching using 'GROW' model[d] including 2-3 suggestions for improvement + expert demonstration followed by learner demonstration of desired improvements, as required + action plan.	✓				✓	✓	✓		✓		0.3 (-0.44 – 1.05)

			(ongoing clinical work as a surgical resident).											
Vafaei 2017 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 x 5min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓					**3.04** (1.95 – 4.13)
Xeroulis 2007 (fb after) Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	Medical students UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. 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 – 1.80)
Xeroulis 2007 (fb during)	Same as above	Same as above	Same as above	Same as above except expert fb during practice trials.	✓		✓			✓	✓				**1.44** (0.43 – 2.46)

Abbreviations:
% = percentage; CI= confidence interval; CPR = cardiopulmonary resuscitation; GI = gastrointestinal; Info = information; M = male; O&G = obstetrics and gynaecology; PGY = postgraduate year; SMD= standardised mean difference; UGY = undergraduate year; wk =week/s

Footnotes:
a= See Figure 5 forest plot for additional study details.
b= Machine output information: simulator metrics (e.g. procedural time or instrument path length) or CPR machine information (e.g. compression rate and depth)
c= Coaching model adapted from Center for Creative Leadership Coaching for Development;
d= Peytons' 4 step model[36]

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',[29, 31] six studies were rated 'unclear'[25, 26, 28, 30, 32, 34] and three studies were 'high risk'.[27, 33, 35] (See 'Risk of bias' section in the supplementary material for more details - Appendix 2).

[Figures 2 and 3 here]

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.

[Figure 4 here]

Meta-analysis

A meta-analysis of the impact of verbal face-to-face feedback compared to 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 provided *during* practice in one group and *directly after* practice in another were compared to no feedback;[32] in another study, feedback provided by an *expert* in one group and by a *peer* in another group[26] were compared to no feedback. In the meta-analysis, numbers for the control group for these studies were halved to retain sample independence.[21]

The meta-analysis of the effect of verbal face-to-face feedback compared to no feedback on workplace task performance found a standardised mean difference of 1.09 (95% confidence

interval (CI) 0.59-1.59; *P*<0. 001) using a random-effects model. The forest plot is presented in Figure 5a.

[Figure 5 here].

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 ten comparisons.[25, 26, 28-32, 34] The standardised mean difference was 0.7 (95% CI 0.37-1.03; *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 moderate, using the GRADE approach. We downgraded the overall rating for the certainty of the evidence from high to moderate, in view of the risk of bias (in particular, due to a lack of concealment and potential for selective reporting of outcomes), methodological and statistical heterogeneity and a high probability of publication bias. (See 'Certainty of evidence' section in the supplementary material for more details - Appendix 2). Figure 6 displays the Summary of Findings table.

[Figure 6 here]

Comparison 2: The effect of verbal face-to-face feedback, compared to alternative feedback, on performance.

Included studies

Table 2 presents the characteristics of included studies in the comparison of verbal face-to-face feedback compared to alternative feedback. Twenty studies (22 comparisons) were included in this analysis and involved verbal, face-to-face feedback compared to alternative feedback. Nine studies (9/20, 45%) were published in the last five 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%) males from 13 studies that reported gender data.[28, 33, 37-47] Included studies involved students (medical, mixed health professions and pharmacy) (1869, 95%) in 16 studies,[28, 32, 33, 37-40, 42-50] and doctors (105, 5%) in four studies.[26, 34, 41, 51] All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[34, 41] Three studies evaluated performance in a clinical practice setting (involving teaching skills,[34] professional and communication skills[41] and oral case presentations[50]) and the remaining 17 assessed performance in a simulated environment (surgical procedures, nasogastric tube insertion, intubation, hearing test, pharmacy consultation or CPR).[26, 28, 32, 33, 37-40, 42-49, 51] (See 'Included studies' section in the supplementary material for more details - Appendix 3).

Risk of bias

In summarising the risk of bias across domains within each study, two studies were rated as low risk,[38, 45] seven studies were rated as 'high risk',[33, 42, 46, 47, 49, 51] 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 - Appendix 3).

Effect of interventions

Figure 7 presents the forest plot and standardised mean differences (SMD). One additional study[38] 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 odds ratio of 1.25 (95% CI: 0.85-1.84) that favoured the learning conversation. The feedback comparisons were markedly diverse, so we did not pool outcomes in meta-analysis.

[Figure 7 here]

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (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 feedback to an individual health professional	**Intervention B**	**Study outcomes**[a] Unless otherwise stated, effects are SMD (95% CI) in favour of feedback intervention A Bolded text indicates significant effects ($p<0.05$)
Al-Jundi[b] **2017** 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 x 10 mins for baseline performance. Performance evaluation: 2 days later	**Immediate face-to-face + written expert feedback** 1 x 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 x 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 – -0.79) significant effects in favour of feedback intervention B

Backstein 2005 Canada	Simulated surgical procedure using a bench top model (vascular anastomosis)	**Doctors in surgical training** PGY1	Intervention duration:4 wk Lecture on surgical procedure. 3 x 2h 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 wk 4	**Review of performance video with expert fb** **+ practice sessions with expert fb available** 3 x weekly videotaping of surgical procedure, with expert feedback available during task, followed by up to 15min review of video with expert fb	**Practice sessions with expert fb available**	**0.86** (0.05 – 1.67)
Baldwin 2015 England	Simulated BLS	**Health professional students** medical (58%), physio (12%), pharmacy (10%), nursing (10%), dentistry (10%) UGY1 33% M	Intervention duration: 4 wk Instruction and practice with manikin 3 x 2.5h 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 wk 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[c] with final summary.	**'Feedback sandwich model'** Fb involved a point for improvement in between 2 points of praise.	**OR 1.25**[f] (95% CI: 0.85-1.84) significant effects in favour of feedback intervention A
Boehler 2006 USA	Simulated surgical skill using a bench top model (tying a 2-handed square knot)	**Medical students** UGY2-3 52% M	Intervention duration: 1 session Instruction in knot tying from surgeon. 1 x baseline performance. Performance evaluation: end of session.	**Expert feedback** 1 x episode of fb from expert surgeon, directly after performance, describing 1-2 specific ways to improve performance.	**Compliment** 1 x pre-scripted general compliment e.g. 'great job!'	**0.98** (0.25 – 1.71)
Bosse 2015 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 4 step procedural training	**High frequency fb** 6 x episodes of fb, given after each practice trial.	**Low frequency practice** 2 x episodes of fb, given after first and last practice trial.	**0.81** (0.21 – 1.40)

			method[d] 6 x practice trials. Fb 'positively worded', focused on effect of actions, given directly after performance by senior peer instructors, trained in procedure & fb. Performance evaluation: end of session.			
Boyle 2011 Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty + stent)	**Doctors training in surgery** PGY4+	Intervention duration: 1 session Teaching & expert demonstration. Fb providers had simulator training. 5 x practice trials (each maximum 40min). Performance evaluation: end of session.	**Expert fb** 5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	**Peer fb** 5 x fb episodes. Peer discussed simulator output, any task errors & the teaching instructions given at start of session.	0.46 (-0.70 – 1.61)
Brinkman 2007 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 feedback 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 feedback as part of clinical training** 1 x 30min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses & parents. Used a coaching model[e] focused on assisting learner to understand information, design goals and improvement strategies. Fb given by paediatricians trained in coaching model.	**Routine feedback as part of clinical training** Performance ratings from nurses and patients not seen.	**2.70** (1.75 – 3.64)
DeLucenay 2017	Simulated pharmacist patient	**Pharmacy students**	Intervention duration: 1 semester.	**Immediate face-to-face fb**	**Delayed written fb**	0.30

USA	consultation (identifying prescription errors and communication skills)	UGY3	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 5min fb on communication skills. Performance evaluation: last 4 SP consultations.	4 x 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.	4 x videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.	(-0.02 – 0.62)
Lee 2016 Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis)	**Medical students** UGY3-4 78% M	Intervention duration: 3 wk Instruction and expert demonstration of procedure, followed by 3 x weekly 30min practice sessions. Performance evaluation: end of 3rd session.	**Early feedback** 1 x 10-15min expert fb directly after first practice attempt, focused on assessment domains.	**Late feedback** Same as early fb but at end of second practice session.	**1.3** (0.26 – 2.34)
Manzone 2014 (verbal comment focused on performance vs verbal comment + comparison to training levels) Canada	Simulated intubation using manikin	**Medical students** UGY1-2	Intervention duration: 1 session Instructional video on intubation. 1-1.5h practice with manikin, with learner in 4 different positions (5 x practice trials in each position). 10 x fb by expert, given directly after practice trials in 2 positions (2 x 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 2 aspects of performance (either done correctly or not) e.g. 'improper use of the laryngoscope'. + individual's progress on task.	**Performance comment compared to others (different training levels)** Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standards at different training levels e.g. 'your performance was at the level of a resident.'	-0.93 (-1.89 – 0.03)

Manzone 2014 (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 – 0.51)
Manzone 2014 (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 to 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 – -1.55) significant effects in favour of feedback intervention B
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk 1st session: 2h instruction and practice suturing & tying knots 'until able to do it easily'. 2nd session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30mins familiarisation with equipment. Practice: 1h daily, 6 days per week for 4 weeks Simulator output information available at the end of each practice session: task completion	**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 & given target goals. .	**Simulator output information**	0.51 (-1.16 – 2.19)

			time, smoothness of tool manipulation and path length of tool.			
Ozcakar 2009 Turkey	Simulated patient consultation with a simulated patient (communication and history taking skills)	**Medical students** UGY2 62% M	Intervention duration: 2 wk Study conducted during routine university module on clinical skills training. Evaluation: 2 wk after intervention following clinical skills lectures + practice with video recording.	**Video review with expert + expert fb** 1 x videotaping of SP consultation. Directly afterwards, review video with expert plus fb.	**Expert fb** 1 x expert fb directly after SP consultation	0.32 (-0.23 – 0.87)
Pavo 2016 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 x 2h additional session including training using modified Peyton 4 step approach[b] 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 & depth, hand position, decompression & 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 & depth plus automated audio advice to correct any deviations during CPR.	-0.09 (-0.36 – 0.18)
Rogers 2012 USA	Simulated surgical skill (tying a single 2-handed square knot)	**Medical students** 'surgical clerkship year'	Intervention duration: 1 session Training in knot tying. 2 x practice trials (1 before & 1 after training). Performance evaluation: end of session.	**Expert fb** 1x fb from expert, with specific information on improving subsequent performance, directly after performance.	**Compliment** 1 x general compliment from expert, instead of fb.	**1.69** (1.06 – 2.32)
Skeff 1983	Clinical teaching skills during ward	**Attending Physicians**	Intervention duration: 1 month	**Expert peer fb** 1 x 1h session mid-term with expert peer, including review of	**Written fb**	-0.36 (-1.06 – 0.34)

USA	round in routine clinical practice		Performance evaluation: medical students and junior medical staff (trainees) on ward rated physicians' teaching skills during ward rounds, at the mid- and end of 1 month term.	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	Received written summary of trainees' evaluation of physician's teaching skills.	
Sox 2014 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 x 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 x constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw 1 item evaluation form but not given a copy.	0.15 (-0.07 – 0.37)
Strandbygaard 2013 Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extra-uterine pregnancy)	**Medical students** UGY 4-6 44% M	Intervention duration: 2 months 1 x 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	**Standardised expert fb with later, additional expert fb if requested by learner + simulator performance score** 1-3 x 10-12min episodes of expert fb involving information on how to perform task components correctly. 1st fb episode provided after first practice trial; learner could ask for up to 2 additional fb episodes (optional) involving same standardised advice + simulator performance score.	**Simulator performance score**	**1.31** (0.86 – 1.77)

			proficiency level; could practice daily (max 3h) for up to 2 months.			
Van de Ridder 2015a (Advances in Health Science Education) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 35% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. Fb from senior medical student with acting experience & 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** 1x fb directly after baseline performance. Fb comment started with global praise followed by the most suitable suggestion for improvement, selected from a list of 4 commonest task errors (e.g. 'You did this well; a tip is …')	**Negatively framed fb** 1x episodic fb directly after practice trial. Fb comment started with global criticism followed by most appropriate directive advice for improvement, selected from list of 4 commonest task errors. (e.g. 'You did not do this correctly; you should change.')	0.41 (-0.06 – 0.87)
Van de Ridder 2015b (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 31% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. All fb providers trained for 1h 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 x fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag & introduced as Professor ENT, wearing a white coat).	**Low credibility fb provider** 1 x fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female, informally dressed).	-0.23 (-0.71 – 0.26)
Xeroulis 2007 Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task.	**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 – 0.73)

			Practice involved 19 x trials in 1h. Fb involved constructive ways to improve + expert demonstration. Performance evaluation: end of session.			

Abbreviations:
% = percentage; BLS = basic life support; CI= confidence interval; CPR = cardiopulmonary resuscitation; ENT = ear, nose and throat specialist; ERC = European Resuscitation Council; fb= feedback; GRS= global rating scale; h = hour; Max = maximum; min = minutes; NG= nasogastric; NR= not reported; UGY = undergraduate year (referring to university year level); physio= physiotherapy; PGY = postgraduate year (referring to post-qualification year); SMD= standardised mean difference; SP = simulated patients; VR = virtual reality; W&R = Weber & Rinne test; wk = week/s

Footnotes:
a= See Figure 7 forest plot for additional study details; b = additional data obtained from authors, enabling calculation of mean, SD and % men; c = Advocacy Inquiry approach[52]; d =Peyton's 4 steps[36]; e= Coaching model adapted from Center for Creative Leadership Coaching for Development. f= Categorical data only available (see text in Results for more details).

DISCUSSION

Comparison 1: The effect of verbal face-to-face feedback, compared to no feedback, on performance.

Our analysis found that verbal face-to-face feedback substantially enhances health professionals' performance compared to no feedback, with SMD 0.7 (95% CI 0.37-1.03; $P<0.001$) from eight studies involving 392 health professionals, after excluding studies at high risk of bias. We are moderately certain about the evidence for this outcome. Nevertheless, this result is constrained by the limitations in the quality of the included studies. To our knowledge, this is the first report to substantiate the widely held view that feedback enhances performance and to estimate the benefit. Included studies involved health professional students and clinicians (mainly medical) performing a range of workplace tasks, particularly surgical and most commonly in a simulated environment. Both results are dominated by one study,[28] evaluating effective compressions during CPR, which contributed the largest number of participants from a single study to the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02-0.51). Several factors may have contributed to the relatively small overall benefit from this feedback intervention compared to many of the other included studies. These include a short practice period with feedback from a peer (as opposed to an expert) who was concurrently performing a different task (the student performing ventilation provided advice on correcting compressions to the student performing compressions).

The consistent positive effects across all included studies supports that the average effect of verbal face-to-face feedback in the health professions is very likely to enhance performance. The pooled effect size was moderately large at 0.7;[21] this indicates that someone at the 50th centile (i.e. mean performance score) in the feedback group would be at the 76th centile performance score in the no feedback group.[53, 54] In comparison a meta-analysis by Kluger and

DeNisi,[55] which analysed feedback (any type across any discipline) compared to no feedback, reported a pooled SMD of 0.4 (notably one third of included studies reported a detrimental impact). To provide more context regarding effect sizes for feedback, the largest beneficial effects of feedback reported in Kluger and DeNisi's meta-analysis were i) effect size 0.55 when feedback included information on any changes since the previous attempt, ii) effect size 0.51 when a specific and challenging goal was set, iii) effect size 0.47 when feedback posed little threat to self-esteem and iv) effect size 0.43 when feedback included information on the correct outcome.

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 likely range in feedback effect for any individual study, from a very small detrimental impact to a very large beneficial effect on performance. This aligns with previous meta-analyses within the health professions and beyond which have reported diverse impacts from different feedback interventions.[14, 18, 19]

Comparison 2: The effect of verbal face-to-face feedback, compared to alternative feedback, on performance.

For the second comparison of the effect of verbal face-to-face feedback compared to 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: additional expert coaching sessions compared to routine monthly written feedback from supervisors;[41] expert feedback early in a practice period compared to later;[43] additional episodes of feedback from experts;[40, 45] additional episodes of feedback involving expert video analysis[51] and expert feedback compared to compliments.[39, 48]

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

Our review focused on the effect of feedback on performance. The studies assembled in this review illustrate the wide variety of possible elements contained within a feedback intervention and the potential influence on performance. Starting with verbal face-to-face feedback interventions in included studies, important differences were seen between studies involving feedback content, source and timing. Detailed specifications about feedback content were often not clearly reported, which suggests that researchers may not have realised the importance of this factor. Previous research has identified that feedback is more effective when it makes the goal clear (for example, describing correct performance or providing an expert demonstration of the task) and advice on how to improve.[10, 19, 55] In our included, the source of feedback was more often experts than peers. One small study[26] directly compared expert feedback to peer feedback for novices learning a surgical task using a visual reality simulator. It did not find a statistically significant difference (SMD 0.46, 95% CI -0.7-1.61), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother. In earlier systematic reviews[18, 19] and other research,[56-60] feedback from a highly credible source (expert feedback) has been reported to be more effective. In the studies included in our review, the timing of feedback also varied; it was provided while the learner undertook the task, immediately afterwards or some time afterwards. One small study, [32] in which novices learnt to suture, feedback during the task was compared to feedback immediately after each attempt. It did not find a statistically significant difference in performance after one hour of practice but did a month later (beyond the scope of our review), in favour of feedback immediately after practice. In another study, in which students practised simulated laparoscopic surgery, the effect of additional expert feedback was compared with performance information provided by the simulator alone. The authors reported that more participants in the 'simulator

feedback only' stopped practising. Previous research has noted that for novices learning a complex task, early feedback and assistance may prevent extreme frustration and giving up.[56] Feedback during task performance results in faster initial skill acquisition compared to feedback after task performance, particularly for procedural skills, as errors are corrected in real-time, but poorer subsequent independent performance.[56, 61] It is thought that a learner develops a mental schema depicting how to do the task, which they develop during practice attempts and this is utilised for subsequent performances.[62, 63] However feedback during task performance appears to interfere with this process, possibly due to cognitive overload.[43, 64]

In the second analysis, verbal face-to-face feedback was compared to a multitude of feedback variations. In addition to feedback (source, frequency, timing) and content, there were differences across feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or 'feedback sandwich'), phrasing of feedback (expressing the same corrective information in a positively or negatively couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into components that might influence feedback outcomes. However, there was insufficient evidence to clearly identify and recommend effective feedback components, as lines of research enquiry had not been systematically pursued with sufficient replication and rigour.

Influences on performance due to factors beyond feedback

Performance improvement is not solely related to feedback. In our review, other important factors influencing performance also varied between studies. Firstly, teaching and expert demonstration were common (but not standard) and the amount and type varied across studies.[36, 65] Practice opportunities also differed enormously across included studies, even those involving similar tasks. In addition, there was variation across learners' prior relevant expertise

(e.g. first year medical students or surgical trainees learning a surgical task) and the complexity of the task (knot tying or laparoscopic bariatric surgery). Previous research has shown that teaching and expert demonstration assist a learner to improve, particularly in the initial phases of skill acquisition[36, 65] and practise is essential for mastering any skill.[15, 66] Furthermore, learners who are motivated to learn a challenging but achievable skill are most likely to improve their performance, according to 'goal setting' and 'self-determination' theories.[10, 67]

Review strengths and limitations

In this systematic review, we have reported for the first time, the benefit of verbal face-to-face feedback, compared to no feedback, on health professionals' task performance. The effect of feedback was positive in all the studies included in the meta-analysis, suggesting a consistent direction of effect. Our review also summarised and categorised multiple variations in feedback interventions (both verbal face-to-face and alternative feedback) and described other factors that have been reported to influence performance in previous research; this clarifies key parameters that need to be considered in future research into feedback.

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).[68]

Implications for practice and future research

Our review fills an important evidence-gap and supports the beneficial impact of verbal face-to-face feedback on health professionals' task performance, compared to no feedback. Our

review also highlights the substantial variability encompassed by the term 'feedback intervention'. By analysing included studies based on key parameters known to influence performance, our review may assist future researchers to refine their methodology and enrich the value of their results by considering these factors. Many of the included studies were 'one-off', involved small numbers of participants and included sources of bias. To advance this field of knowledge, research programs designed to systematically investigate the feedback components required for effective feedback are needed. This is likely to involve a series of studies designed to isolate one feedback component at a time, with all other key influences on performance standardised, in order to identify, replicate and validate the conditions that are most effective in helping learners to improve, across different contexts.

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 substantially enhanced workplace task performance compared to no feedback SMD 0.7 (95% CI: 0.37-1.03; $P<0.001$), 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, incorporating multiple distinct components. More robust evidence is required and future research should focus on systematically analysing components that maximise the effects of feedback.

Declarations of interest: The authors report no declarations of interest.

Author Contributions: CJ and JK designed the review, CJ created the search strategy and undertook the literature searches, CJ and MW undertook study screening, data extraction and quality assessment with oversight from JK, CJ and JK undertook data analysis and interpreted the findings, CJ prepared the manuscript and all authors contributed to the final version.

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

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: The authors report no competing interests.

Patient consent: Not required.

Ethics Approval: Not required.

Data sharing statement: Data access can be requested by contacting the corresponding author.

Figure and Table Legends

Figure 1: PRISMA flow diagram.

Figure 2: Risk of bias graph

Figure 3: Risk of bias summary

Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 6: Summary of findings table for the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance.

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

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

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(5):645-9. doi: 10.1097/acm.0000000000001048 [published Online First: 2015/12/18]
3. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34(3):205-14. doi: 10.3109/0142159x.2012.652239 [published Online First: 2012/03/01]
4. Holmboe ES. Realizing the promise of competency-based medical education. *Academic Medicine* 2015;90(4):411-3. doi: 10.1097/acm.0000000000000515 [published Online First: 2014/10/09]
5. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association* 2009;302(12):1316-26. doi: https://dx.doi.org/10.1001/jama.2009.1365
6. Archer JC. State of the science in health professional education: effective feedback. *Medical Education* 2010;44:101-08. doi: 10.1111/j.1365-2923.2009.03546.x
7. Bing-You RG, Hayes V, Varaklis K, et al. Feedback for learners in medical education: what is known? A scoping review. *Academic Medicine* 2017;92(9):1346-54. doi: https://dx.doi.org/10.1097/ACM.0000000000001578
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 Medical Education* 2016;16(1):96. doi: 10.1186/s12909-016-0613-5
9. Watling C, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ* 2019;53(1):76-85. doi: 10.1111/medu.13645 [published Online First: 2018/08/04]

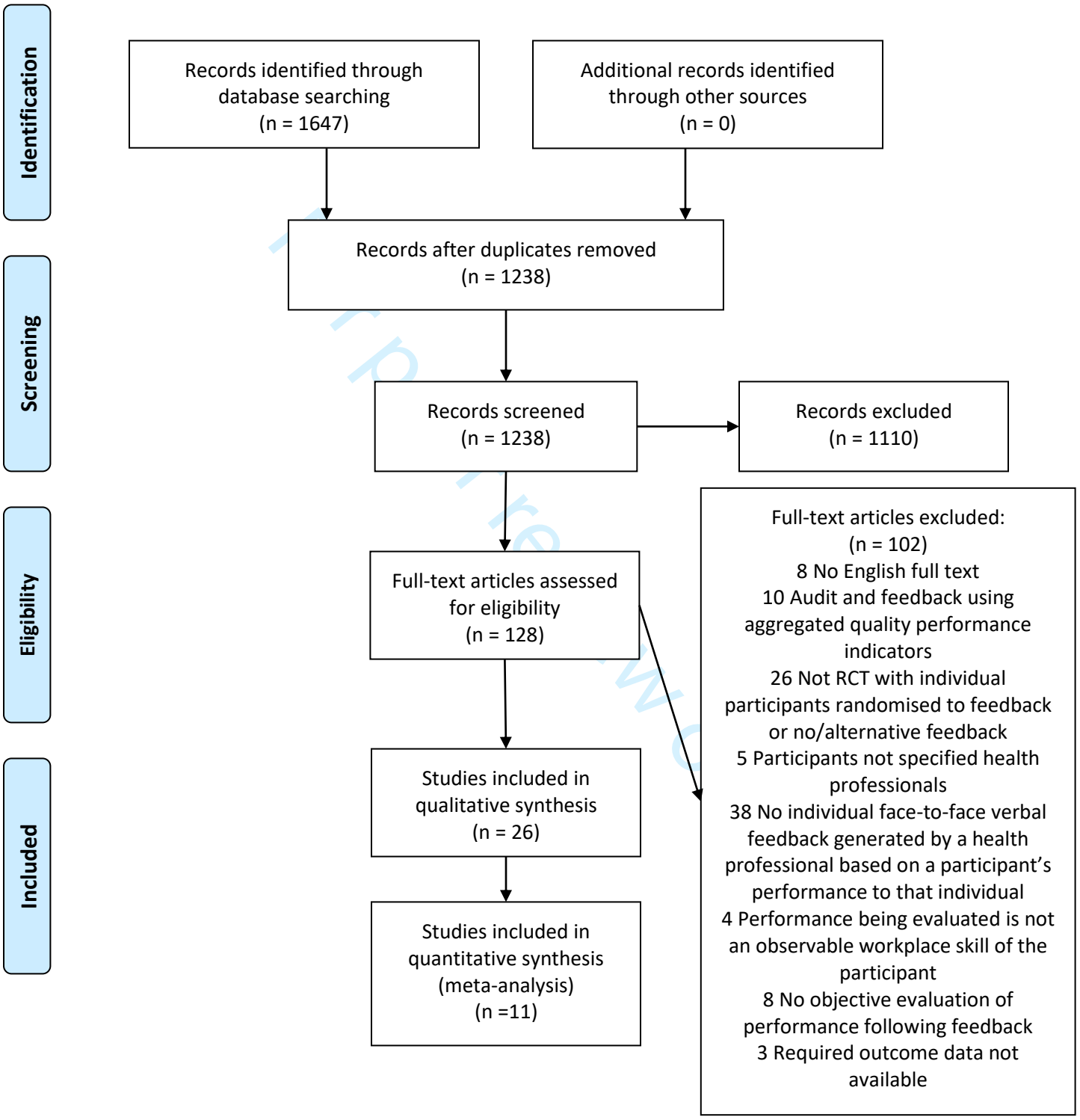
10. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist* 2002;57(9):705-17. [published Online First: 2002/09/20]
11. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine* 2015;90(5):609-14. doi: https://dx.doi.org/10.1097/ACM.0000000000000560
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. *Medical Teacher* 2011;33(12):961-73. doi: https://dx.doi.org/10.3109/0142159X.2011.595435
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 Medical Education* 2019;19(1):129. doi: 10.1186/s12909-019-1524-z
14. Hattie J, Timperley H. The Power of Feedback. *Review of Educational Research* 2007;77(1):81-112. doi: 10.3102/003465430298487
15. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
16. Molloy E, Boud D. Changing conceptions of feedback. In: Boud D. ME, ed. Feedback in higher and professional education London: Routledge 2013:11-33.
17. Ende J. Feedback in clinical medical education. *Journal of the American Medical Association* 1983;250(6):777-81. doi: 10.1001/jama.1983.03340060055026
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. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
19. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
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(4):264-9, w64. doi: 10.7326/0003-4819-151-4-200908180-00135 [published Online First: 2009/07/23]
21. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration 2011 [Available from: www.handbook.cochrane.org.
22. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
23. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List. *Journal of Clinical Epidemiology* 1998;51(12):1235-41. doi: 10.1016/S0895-4356(98)00131-0
24. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal* 2011;342:d549.
25. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
26. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
27. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
28. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9

29. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
30. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
31. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
32. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
33. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
34. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
35. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
36. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
37. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
38. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
39. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
40. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5
41. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine* 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44
42. 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(1):322-28.
43. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
44. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
45. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e

46. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
47. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
48. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009
49. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397
50. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209
51. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
52. 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 Medical Simulation* 2016;1(1):49-55.
53. Sullivan GM, Feinn R. Using Effect Size -or Why the P Value Is Not Enough. *Journal of Graduate Medical Education* 2012;4(3):279-82. doi: 10.4300/JGME-D-12-00156.1
54. Coe R. It's the effect size, stupid: What effect size is and why it is important. A paper presented at the Annual Conference of the British Educational Research Association 2002. 2002
55. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
56. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
57. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
58. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
59. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.
60. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
61. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360
62. Elliott SN. Educational psychology: effective teaching, effective learning. 3rd ed. Boston: McGraw-Hill 2000.
63. Wadsworth BJ. Piaget's theory of cognitive and affective development: foundations of constructivism., 5th edn. White Plains: Longman Publishing 1996.
64. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
65. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.

66. 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 ed. Oxford: Wiley Blackwell 2014:7-29.
67. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227-68. doi: 10.1207/s15327965pli1104_01
68. Belfield C, Thomas H, Bullock A, et al. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;23(2):164-70. doi: 10.1080/01421500020031084 [published Online First: 2001/05/24]

Figure 1: PRISMA flow diagram for systematic review of verbal face-to-face feedback compared to no or alternative feedback.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

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

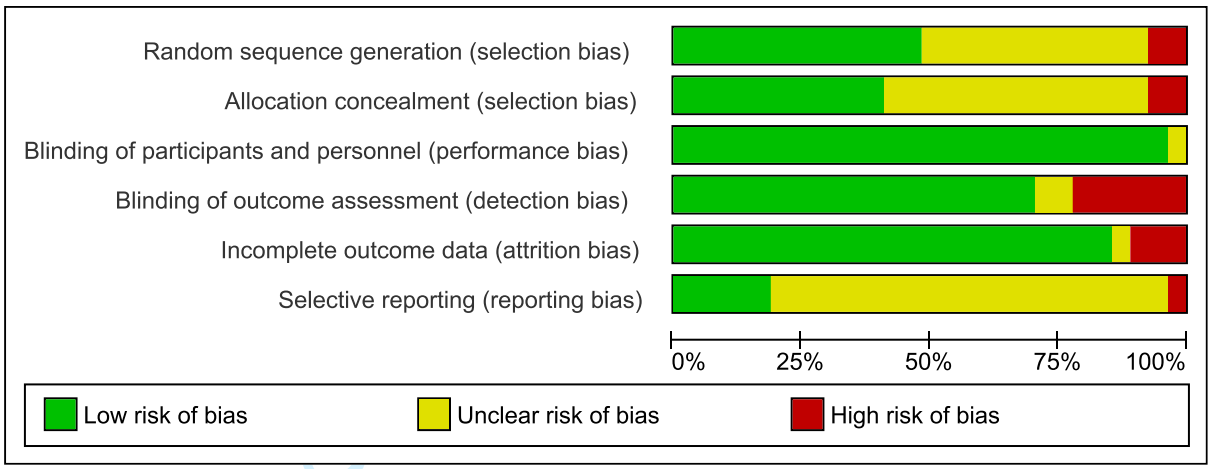


Figure 3: Risk of bias summary: review authors' judgements about each risk of bias item for each included study

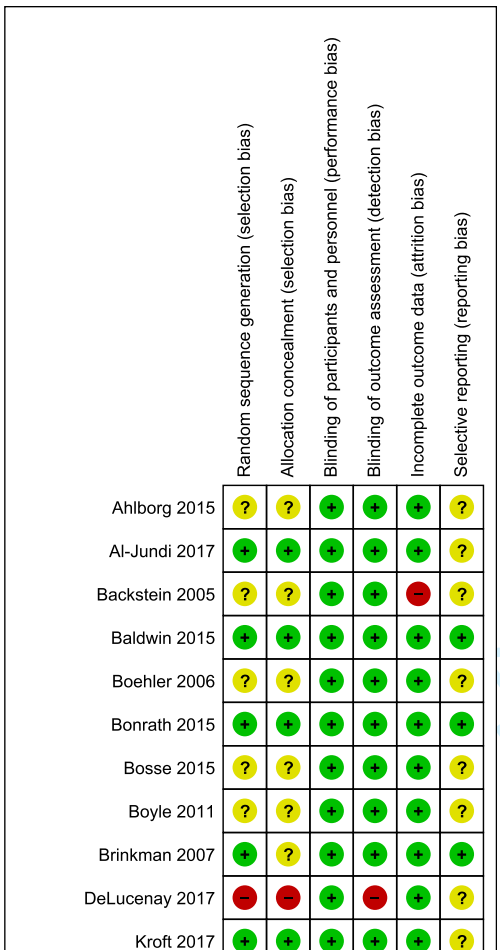
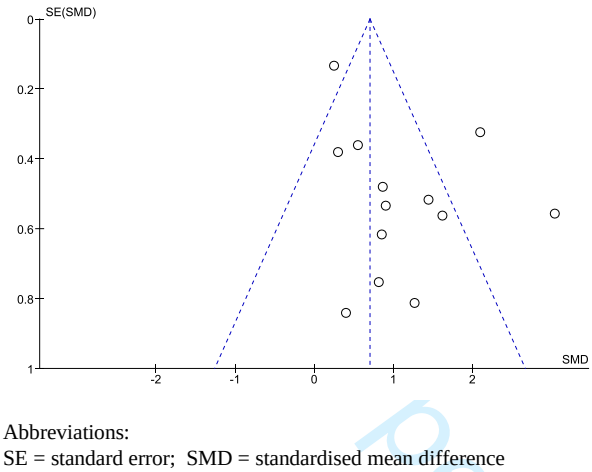
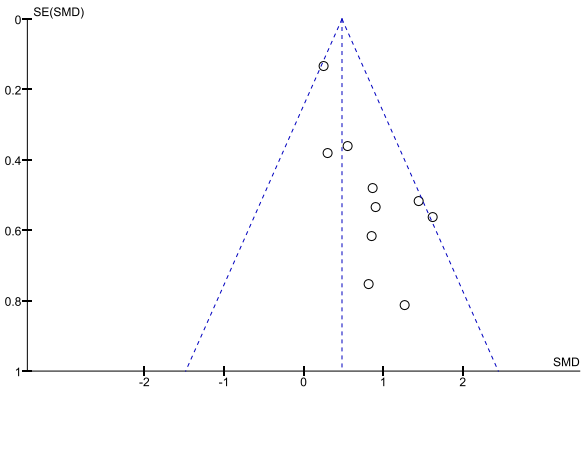


Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

a) all included studies



b) excluding studies at high risk of bias



Abbreviations:
SE = standard error; SMD = standardised mean difference

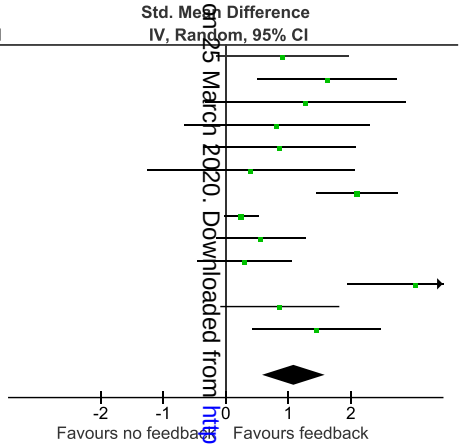
Footnote:
Meta-analysis calcuated using a fixed effects model.
The dotted vertical line represents the overall effect estimate and the dotted slanted lines represent the 95% confidence interval.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance

a) All included studies

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.6%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.3%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	5.3%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	5.7%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.9%	0.85 [-0.35, 2.06]
O'Connor 2008	-1.96	1.25	3	-2.5	0.9	3	5.1%	0.40 [-1.25, 2.04]
Olms 2016	-1.1	1.8	30	-7.4	3.8	30	9.6%	2.09 [1.45, 2.73]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	11.0%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	9.2%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	9.1%	0.30 [-0.44, 1.05]
Vafei 2017	17.73	1.09	15	12.13	2.29	15	7.4%	3.04 [1.95, 4.13]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	8.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	7.7%	1.44 [0.43, 2.46]
Total (95% CI)			**253**			**235**	**100.0%**	**1.09 [0.59, 1.59]**

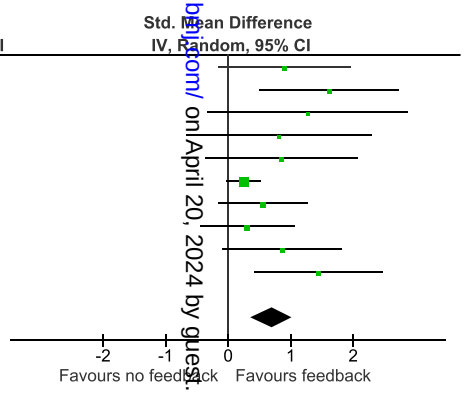
Heterogeneity: Tau² = 0.58; Chi² = 54.26, df = 12 (P < 0.00001); I² = 78%
Test for overall effect: Z = 4.26 (P < 0.0001)



b) Excluding studies at high risk of bias (sensitivity analysis)

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.7%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.1%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	3.8%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	4.4%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.2%	0.85 [-0.35, 2.06]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	27.9%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	13.3%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	12.4%	0.30 [-0.44, 1.05]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	9.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	8.1%	1.44 [0.43, 2.46]
Total (95% CI)			**205**			**187**	**100.0%**	**0.70 [0.37, 1.03]**

Heterogeneity: Tau² = 0.08; Chi² = 13.64, df = 9 (P = 0.14); I² = 34%
Test for overall effect: Z = 4.14 (P < 0.0001)



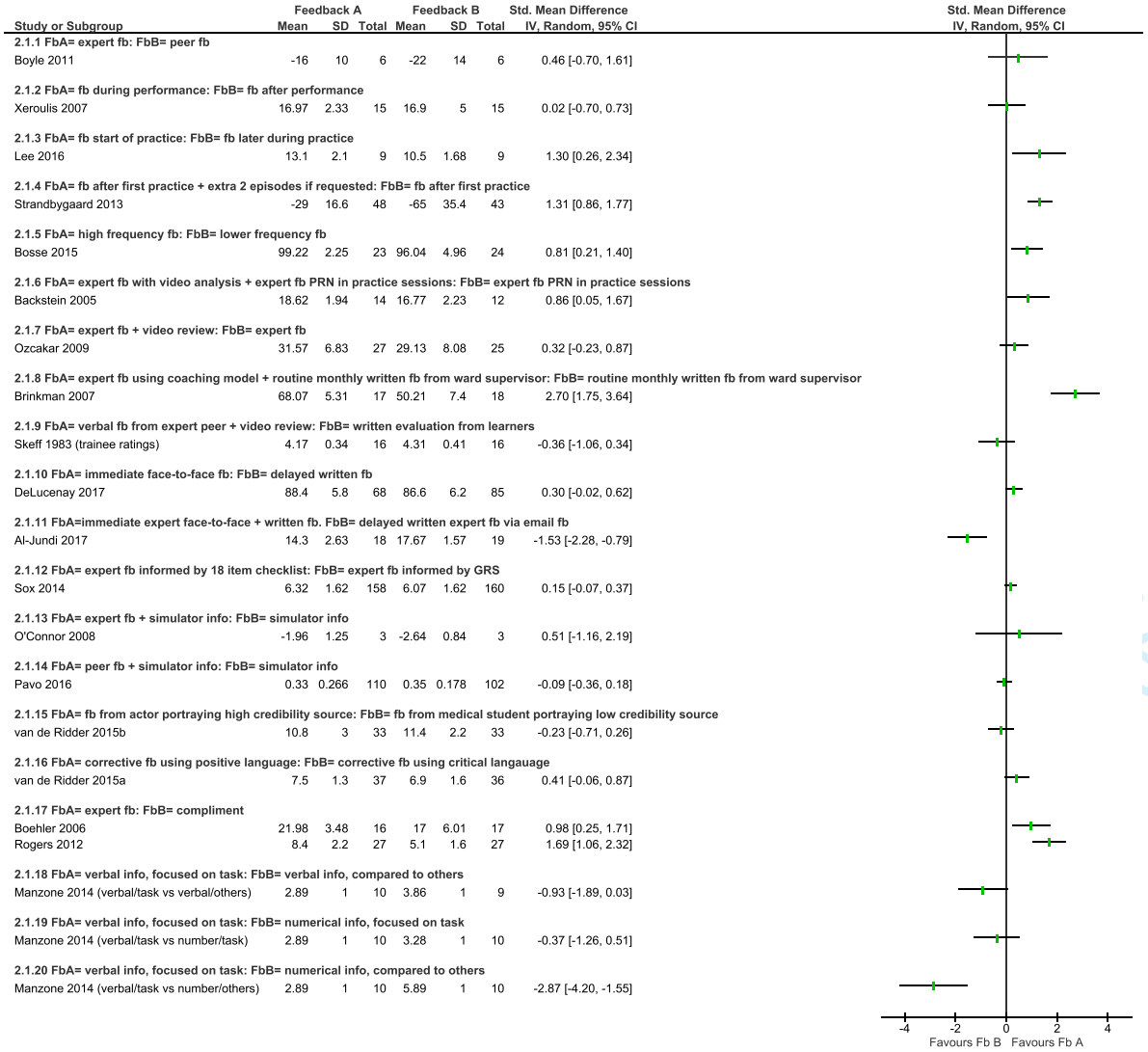
Abbreviations:
SD = standard deviation; CI = confidence interval; SMD = standardised mean difference

Footnotes:
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

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

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 (0.37-1.03) higher than mean scores for the control group	Number of participants 392 (8 studies)	⊕⊕⊕O[a,b,c] Moderate	A SMD of 0.7 indicates a substantial improvement in performance
CI = Confidence interval; SMD= standardised mean difference				
[a]High risk of bias due to lack of allocation concealment and prior published protocols to counter selective reporting of outcomes. [b]Unexplained heterogeneity [c]High probability of publication bias				
GRADE Working Group grades of evidence ⊕⊕⊕⊕ High: We are very confident that the true effect lies close to that of the estimate of the effect. ⊕⊕⊕O Moderate: We are moderately confident in the effect estimate, the true effect is likely to be close to this but there is a possibility that it is substantially different ⊕⊕OO Low: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect ⊕OOO Very Low: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.				

Figure 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance



Abbreviations:
Fb= feedback; GRS= global rating scale; info= information; PRN= 'as required'; SD = standard deviation; CI=confidence interval

Footnotes:
[1]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

bmjopen-2019-030672 on 25 March 2020. Downloaded from

Database(s): Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:

#	Searches	Results
1	*Feedback/	6031
2	Feedback, Psychological/	3311
3	Formative Feedback/	467
4	(feedback adj3 (effective or formative or constructive or quality or clinical or performance)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4860
5	1 or 2 or 3 or 4	13233
6	exp Health Personnel/	470058
7	exp Health Occupations/	1648689
8	exp Dentistry/	386159
9	exp Social Work/	17331
10	exp Psychology/	66579
11	Occupational Therapy/	13213
12	Radiotherapy/	42757
13	Radiography/	334082
14	Mentors/	9949
15	exp Students, Health Occupations/	60760
16	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2722708
17	clinician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	171551
18	(health* adj2 (staff or personnel or faculty or provider* or worker* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	303659
19	doctor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104240
20	physician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	495037
21	(medical adj3 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	133207
22	general practitioner*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45015
23	(general pract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45488
24	(family adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21549
25	(primary care adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10661
26	(primary health* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2156
27	(registrar or registrars or senior house officer* or resident or residents or hospital medical officer* or intern or interns or house officer*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	129452
28	dentist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	115221
29	(dent* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21015
30	nurs*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	629304
31	(midwife or midwives).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	18155

32	(midwife* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	965
33	(allied health adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	12936
34	physiotherapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5348
35	physical therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5123
36	(physiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	731
37	(physical therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	647
38	occupational therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4500
39	(occupational therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1058
40	speech therap*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7630
41	(speech therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	22
42	speech language therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	136
43	(speech language therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	14
44	speech pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	656
45	(speech pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	31
46	speech language pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1561
47	(speech language pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	105
48	dietician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1352
49	(dietetic* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	603
50	podiatrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	626
51	(podiatry adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	19
52	chiropodist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	119
53	(chiropody adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2
54	social worker*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8079
55	(social work* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8517

56	psychologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	11986
57	(psychology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1532
58	(osteopath* or osteopathic physician*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	6480
59	(osteopath* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	296
60	chiropractor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1148
61	(chiropract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	241
62	pharmacist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	27316
63	(pharmac* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7184
64	optometrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1764
65	(optometr* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	290
66	(Radiographer* or radiological technologist* or radiation therapist* or radiotherapist* or radiation therapy technologist*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2449
67	(radiograph* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	267
68	(radiation therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	53
69	(radiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	100
70	(supervisor* or tutor* or trainer* or educator* or teacher* or mentor* or preceptor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104110
71	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70	1846616
72	16 or 71	3625236
73	exp Education, Professional/	284462
74	exp Educational Measurement/	137138
75	exp Professional Practice/	247602
76	exp Simulation Training/	6239
77	(effect* or evaluat* or outcome* or assess* or measur*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10754772
78	73 or 74 or 75 or 76 or 77	11134144
79	randomized controlled trial.pt.	505181
80	controlled clinical trial.pt.	100406
81	randomized.ab.	391590
82	randomly.ab.	266043
83	systematic review.ab,ti.	85419
84	79 or 80 or 81 or 82 or 83	966230
85	5 and 72 and 78 and 84	821
86	limit 85 to (english language and humans)	809

Supplementary material. Appendix 2

RESULTS: SUPPLEMENTARY INFORMATION

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

Included studies

Participants

Participants included 290 (60%) medical students in four studies,[1-4] 60 (12%) dental students in one study[5] and 138 (28%) doctors (doctors training in surgery in three studies,[6-8] training in obstetrics and gynaecology in one study[9] and training in emergency medicine in one study,[10] and physicians in one study[11]).

Participants were novices to the assessed task in five studies (5/11, 45%);[1, 2, 4, 5, 7] and had prior experience in six studies.[3, 6, 8-11]

Workplace tasks and Settings

All studies evaluated performance of a discrete task; there were no longitudinal evaluations. The task occurred in simulation settings in seven studies (7/11, 64%) and clinical practice in four studies (4/11, 36%). The task was a surgical procedure in seven studies (7/11, 64%). Five studies involved simulated surgical tasks including bench top models for knot tying[4] and forming a bowel anastomosis;[8] using a laparoscopic simulator for suturing and knot tying;[2] and using a virtual reality (VR) simulator for laparoscopic surgery[1] and endovascular surgery.[7] Two studies involved laparoscopic surgery in clinical practice.[6, 9] The remaining four studies evaluated simulated matching of tooth colour in a dental school,[5] simulated cardiopulmonary resuscitation (CPR),[3] chest ultrasound for emergency trauma patients[10] and teaching skills in clinical practice.[11]

Feedback Interventions

The feedback source involved a subject expert in all comparisons except two, including one that compared peer feedback with no feedback,[3] and one that compared expert feedback, peer feedback and no feedback.[7] Feedback occurred while the participant performed the task (during) in one study,[3] both during and directly afterwards in two studies,[1, 2] directly afterwards in four studies,[5, 7, 9, 10] after a delay in three studies[6, 8, 11] and one study compared feedback during, feedback directly afterwards and no feedback.[4] In addition to evaluative performance information (as per inclusion criteria), the feedback included corrective advice in all studies except one[10] and one where it was unclear.[7] Feedback included additional information from a simulator in three studies,[1, 2, 7] a video of the participant's performance in two studies[6, 11] and written performance information in two studies.[5, 11]

Teaching and Practice

In addition, instruction and expert demonstration of the task were provided in six studies (6/11, 55%), including all five studies involving novice participants[1, 2, 4, 5, 7] and one study that involved CPR for medical students, many of whom had previously attended a course.[3] The other five studies involved doctors working in clinical practice; in these studies, no instruction or expert demonstration was included within the research intervention but may or may not have occurred during the course of routine work during that time. One study involved physicians' teaching on ward rounds[11] and the other four studies assessed tasks by doctors training in relevant specialties.[6, 8-10]

The amount of practice varied substantially between different studies, for both simple and complex tasks. For example, comparing two studies that involved simple surgical knot tying: in Xeroulis,[4] participants had 18 practice attempts in one session and in O'Connor,[2] they could practice up to an hour a day, for 24 days. Looking at more complex surgical procedures, such as simulated surgery using a virtual reality (VR) simulator: in Ahlborg,[1] participants had two

practice attempts at the simulated surgery (laparoscopic salpingectomy) and in Boyle,[7] participants had five attempts at the simulated surgery (renal artery angioplasty and stenting) before the performance evaluation.

Intervention period

The intervention period ranged from one day (most common) up to two months.[6] Nine (9/11, 82%) studies involved a single session (involving one episode of feedback in five studies[5, 8-11] and multiple episodes of feedback in four studies[1, 3, 4, 7]). Two studies (2/11, 18%) had a longer intervention period involving multiple feedback sessions: one study[6] included approximately four coaching sessions regarding bariatric surgery across a two month surgical attachment, and another[2] included almost daily one hour practice sessions for laparoscopic suturing, with feedback throughout each one, over four weeks.

The timing of the post-feedback performance assessment, in relation to the intervention, differed. It occurred directly following the intervention in seven studies: at the end of the single session in five studies[1, 3, 4, 7, 9] and at the end of an extended intervention period in two studies.[2, 6] In the other four studies, the post-feedback performance assessment occurred some weeks after the intervention was completed but while relevant exposure to possible teaching and/or practice opportunities continued. Olms[5] included a single feedback session, with the final evaluation two weeks later, in the midst of a routine one month university teaching unit on tooth shade matching. Skeff[11] arranged a single coaching session on ward round teaching in the middle of physicians' four week ward duty, with the final evaluation post-performance evaluation at the end. Soucisse[8] also organised a single coaching session for surgical residents, with the final evaluation occurring three weeks later. Vafaei[10] involved a single workplace-based assessment with feedback for doctors training in emergency medicine on chest ultrasound for emergency trauma patients, followed by a two month period of routine clinical work before the post-feedback assessment.

Research funding

Regarding research funding, one study[3] that focused on cardiopulmonary resuscitation (CPR) quality, was loaned a device (used to measure CPR parameters and provide automated feedback to participants) for the period of the study by Philips but the company was not otherwise involved in the research; five studies received funding from independent institutions,[1, 4, 6, 9, 11] three studies did not receive any funding[5, 7, 10] and two studies did not report information on funding.[2, 8]

Risk of bias

Five trials described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[3, 5, 6, 8, 9] The other six trials simply stated participants were 'randomised' and had no information on allocation concealment, so we rated these studies as 'unclear'. We analysed baseline performance because, although randomisation removes the need to check comparability in baseline task performance for intervention and comparison groups, it may be useful to check this when participant numbers are small and performance improvement is more likely when baseline performance is low.[12] Seven studies reported no statistically significant differences between baseline performances for the comparison groups.[4, 5, 8-11] and four studies did not report baseline task performance.[1-3, 7] The participants and research team members were not blinded in any included studies because the intervention involved feedback between a research team member and a participant, consistent with most education interventions. However, in all included studies, we thought this was not likely to influence the outcome (post-intervention performance assessment) because implementation and adherence to the intervention were not affected. In eight studies the outcome was assessed by either blinded assessors who rated videos of the participants' performance[4, 6-9, 11] or by a machine (simulator or CPR machine),[1, 3] so we rated these as 'low risk' of bias. In three studies, the feedback provider and outcome assessor appeared to be the

same person, so these were rated as 'high risk'.[2, 5, 10] Across all the studies, the follow up rate for each group was at least 85%. Only two studies had a prior published protocol in addition to reporting all outcomes as planned.[6, 8] For all other studies, it could not be ascertained if outcomes had been selectively reported, so these were rated as 'unclear', except one. This one study was rated as 'high risk' for selective outcome reporting because it did not include the expected information on performance post-intervention.[2]

In summarising the risk of bias across domains within each study, two studies had all domains rated 'low risk, so these were rated low risk.[6, 8] Six studies had at least one domain with 'unclear' risk but no 'high risk' ratings, so these were rated as 'unclear' risk of bias.[1, 3, 4, 7, 9, 11] Three studies had at least one domain at high risk of bias, so we judged these studies to be at 'high risk' of bias.[2, 5, 10]

Certainty of evidence

For the comparison of verbal face-to-face feedback compared to no feedback, excluding studies at high risk of bias, we graded the quality of evidence for the outcome of 'objective assessment of a health professional's performance'. The risk of bias was rated as 'unclear' across multiple included studies and the overall body of evidence indicated this was likely to seriously alter the results, so we downgraded the overall evidence by one level. The two aspects that were most influential on our decision were the lack of allocation concealment and prior published protocols to preclude selective reporting of outcomes. Participant and research team member blinding was not possible due to the intervention. However, this had limited impact on the selected outcome 'objective assessment of performance', as no changes occurred in intervention implementation or adherence as a consequence of this lack of blinding.[13] We judged the results to be directly applicable to our review question and therefore the evidence was not downgraded for indirectness. There was some methodological and statistical heterogeneity across studies (the test for heterogeneity was not significant with $P = 0.14$ and I^2

= 34%), which was not explained by subgroup analysis. However, all studies reported a beneficial effect, so the uncertainty seemed to lay in the magnitude of effect rather than the presence of an effect. Therefore, we decided not to downgrade the evidence due to inconsistency.[14] We judged the effect size to be sufficiently precise and therefore did not downgrade the evidence for imprecision of results. This was based on sufficient numbers of participants (392 when studies with high risk of bias were excluded) and a consistent beneficial effect, indicated by the confidence interval for the overall effect estimate not crossing zero and all individual studies showing a beneficial effect with substantial overlap in their confidence intervals. Finally, we judged that there was likely to be a systematic overestimation of the underlying beneficial effect of feedback because we strongly suspected publication bias (see Funnel plot 5b) and therefore we downgraded the evidence by one level.

In summary, combining all five GRADE criteria for assessing the certainty of evidence, we downgraded the overall rating by one, from high to low. We judged that the quality of the evidence was low contributing to the effect estimate of 0.70 in the comparison of verbal face-to-face feedback to no feedback after excluding studies with a high risk of bias. Hence face-to-face feedback may result in a moderate to large improvement in health professionals' workplace task performance.

References

1. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
2. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
3. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
4. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012

5. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
6. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
7. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
8. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
9. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
10. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
11. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
12. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
13. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *British Medical Journal* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
14. 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. *Journal of Clinical Epidemiology* 2013;66(2):151-57. doi: 10.1016/j.jclinepi.2012.01.006

Supplementary material. Appendix 4

DISCUSSION

Comparison 1: The effect of verbal face-to-face feedback, compared to no feedback, on performance: supplementary information

Included studies involved health professional students and clinicians (mainly medical) performing a range of workplace tasks, particularly surgical and most commonly in a simulated environment. The meta-analysis results are dominated by one study,[1] evaluating effective compressions during CPR, which contributed the largest number of participants from a single study to the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02, 0.51). Several factors may have contributed to the relatively small overall benefit from this feedback intervention compared to many of the other included studies. These include a short practice period with feedback from a peer (as opposed to an expert) who was concurrently performing a different task (the student performing ventilation provided advice on correcting compressions to the student performing compressions).

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

Previous research has identified that feedback is more effective when the content includes information that makes the goal clear (for example, describing correct performance or providing an expert demonstration of the task) and advice on how to improve.[2-4] However, detailed specifications about feedback content were often not clearly reported in included studies, which suggests that researchers may not have realised the importance of this. The feedback source was more often experts than peers, in our included studies. One small study[5] directly compared expert feedback to peer feedback for novices learning a surgical task using a visual reality simulator. It did not find a statistically significant difference (SMD 0.46, 95%

CI -0.7, 1.61), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother. In earlier systematic reviews[2, 6] and other research,[7-11] feedback from a highly credible source (expert feedback) has been reported to be more effective. Also, the timing of feedback in included studies varied; it was provided while the learner undertook the task, immediately afterwards or some time afterwards. One small study,[12] in which novices learnt to suture, feedback during the task was compared to feedback immediately after each attempt. It did not find a statistically significant difference in performance after one hour of practice but did a month later (beyond the scope of our review), in favour of feedback immediately after practice. In another study, in which students practised simulated laparoscopic surgery, the effect of additional expert feedback was compared with performance information provided by the simulator alone. The authors reported that more participants in the 'simulator feedback only' stopped practising. Previous research has noted that for novices learning a complex task, early feedback and assistance may prevent extreme frustration and giving up.[7] Feedback during task performance results in faster initial skill acquisition compared to feedback after task performance, particularly for procedural skills, as errors are corrected in real-time, but poorer subsequent independent performance.[7, 13] It is thought that a learner develops a mental schema depicting how to do the task, which they develop during practice attempts and this is utilised for subsequent performances.[14, 15] However feedback during task performance appears to interfere with this process, possibly due to cognitive overload.[16, 17]

In the second analysis, verbal face-to-face feedback was compared to a multitude of alternative feedback interventions. In addition to feedback source, frequency, timing and content, there were differences across feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or 'feedback sandwich'), phrasing of feedback (expressing the same corrective information in a positively or negatively

couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into components that might influence feedback outcomes.

Influences on performance due to factors beyond feedback

Firstly, teaching and expert demonstration were common (but not standard) and the amount and type varied across studies, which have previously been shown to impact performance.[18, 19] Practice opportunities also differed enormously across included studies, even those involving similar tasks. In addition, there was variation across learners' prior relevant expertise (e.g. first year medical students or surgical trainees learning a surgical task) and the complexity of the task (knot tying or laparoscopic bariatric surgery). Previous research has shown that teaching and expert demonstration assist a learner to improve, particularly in the initial phases of skill acquisition[18, 19] and practise is essential for mastering any skill.[20, 21] Furthermore, learners who are motivated to learn a challenging but achievable skill are most likely to improve their performance, according to 'goal setting' and 'self-determination' theories.[4, 22]

REFERENCES

1. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
2. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
3. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
4. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist* 2002;57(9):705-17. [published Online First: 2002/09/20]
5. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058

6. Veloski J, Boex JR, Grasberger MJ, et al. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
7. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
8. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
9. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
10. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.
11. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
12. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
13. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360
14. Elliott SN. Educational psychology: effective teaching, effective learning. 3rd ed. Boston: McGraw-Hill 2000.
15. Wadsworth BJ. Piaget's theory of cognitive and affective development: Foundations of constructivism. 5th ed. White Plains: Longman Publishing 1996.
16. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
17. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
18. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
19. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.
20. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
21. 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 ed. Oxford: Wiley Blackwell 2014:7-29.
22. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227-68. doi: 10.1207/s15327965pli1104_01

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	P1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	P2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	P5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	P7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	P7
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	P8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	P8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	P7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	P9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	P9, Tables 1 & 2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	P9-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	P10

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	P10-11

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	P9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	P10-11
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	P11 & Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	P12 & 19-20 Tables 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	P18, & 20 & Figures 2 & 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures 5 & 7
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	P18-19 Figure 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Figures 2 & 3
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	P30-31
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	P3, 34

PRISMA 2009 Checklist

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	P31
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data; role of funders for the systematic review.	P36

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030672.R3
Article Type:	Original research
Date Submitted by the Author:	30-Jan-2020
Complete List of Authors:	Johnson, Christina; Monash Health, Monash Doctors Education ; University of Melbourne, Department of Medical Education, Melbourne Medical School Weerasuria, Mihiri; Monash Health Keating, Jennifer; Monash University, Department of Physiotherapy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	MEDICAL EDUCATION & TRAINING, feedback, health professions education, formative feedback

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Short title: Systematic review and meta-analysis on feedback

Christina E. Johnson MBChB, MHPE, FRACP

ORCID ID: 0000-0002-4209-8419

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, Australia.

Mihiri P. Weerasuria MBBS (Hons)

ORCID ID: 0000-0001-6897-3982.

Monash Health, Melbourne, Australia.

Jennifer L. Keating PhD, GradDipManip PT, BAppSci (PT)

ORCID ID: 0000-0003-3161-4964

Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia.

Correspondence should be addressed to Christina Johnson, Monash Doctors Education, McCulloch House, Monash Health, Clayton Road, Clayton, Victoria 3168, Australia. Telephone: 613-9594 3743. Email: 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 1st 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 to no feedback or alternative feedback, and available as full text publications in English. The certainty of evidence was assessed using the GRADE approach. For feedback compared to 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 to 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 to alternative feedback, studies could not be pooled due to substantial intervention and design 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 to 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.

PROSPERO 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 to 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.

Keywords: feedback, effective feedback, formative feedback, systematic review, meta-analysis, health professions education

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 (for example, from a simulator or within an online learning module). The unique potential benefits of face-to-face verbal feedback are the opportunities for i) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to co-construct insights and strategies for improvement and ii) inter-personal 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 said, "Without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all."[17] 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 which investigated the impact of audit and feedback. In 2006, Veloski et al 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.[18] 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 commonest data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers et al updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared to usual care.[19] 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% (interquartile range 0.5 to 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 to 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 e.g. 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) Registration ID CRD42017081796.

Eligibility criteria for considering studies for this review

We included randomised controlled trials in which individual health professionals were randomised to feedback, compared to 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 pre-determined 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 i) verbal face-to-face feedback compared with no feedback and ii) verbal face-to-face feedback compared to alternative feedback.

Performance following feedback interventions had to be objectively assessed. To isolate the effects of feedback, 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 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 1st February 2019. We also searched the reference lists of systematic reviews and included studies.

Data collection and analysis

Selection of studies

One review author (CJ or MW) screened titles to exclude clearly irrelevant reports. Two authors (CJ and MW) independently screened remaining abstracts to identify potential eligible studies, then independently assessed the full text. Decisions were compared using Covidence (on-line 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 (JK).

Data extraction and management

One review author (CJ) used a pre-piloted standardised form to extract data from included studies and another author (MW or JK) 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.

Assessment of risk of bias in included studies

The risk of bias was independently assessed by two authors (CJ and JK) 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 (for example means (standard deviations SD) or medians (interquartile range IQR)). The effect was quantified using the standardised mean difference 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 standardised mean difference (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 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.

[Figure 1 here]

Comparison 1: The effect of verbal face-to-face feedback, compared to 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 to no feedback on the objective evaluation of a workplace task. Seven (64%) reports were published in the last five 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%) males 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 - Appendix 2).

Table 1: Summary of available data on characteristics of trials included in the comparison of verbal face-to-face feedback (intervention) compared to 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** Additional information	Source		Timing			Content					**Study outcomes**[a] All effects are SMD (95% CI) P value in favour of feedback
					Subject Expert	Peer	During task	Directly after		Verbal performance info	Verbal corrective advice	Machine output info[b]	Performance video	Written performance info	
Ahlborg 2015 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 x practice trials. Performance evaluation: end of session.	2 x fb episodes. Fb given by expert i) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus ii) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓			0.91 (-0.14, 1.95) P = 0.08
Bonrath 2015 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.) x 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, 2.72) P = 0.002

Boyle 2011 (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 x practice trials Performance evaluation: end of session	5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' + simulator output information.	✓			✓		?	?	✓			1.27 (-0.32, 2.87) P = 0.08
Boyle 2011 (peer fb)	Same as above	Same as above	Same as above	5 x fb episodes. Peer discussed simulator output, any task errors & teaching instructions given at start.		✓		✓		✓	✓	✓			0.81 (-0.66, 2.29) P = 0.24
Kroft 2017 Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy)	Doctors training in O&G PGY2-6 33% M	Intervention duration: 1 x 15min practice using laparoscopic salpingectomy module on VR surgical simulator within 1h of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1 x fb episode from expert directly after VR simulator practice. Fb 'standardized and given in an evidence based fashion to optimise effectiveness' & included '3 constructive recommendations based on performance'.	✓			✓		✓	✓				0.85 (-0.35, 2.06) P = 0.14
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	Medical students UGY1-2 44% M	Intervention duration: 4 wk. 2h instruction + practice suturing & knot tying until able to do it easily. Then instruction on laparoscopic surgery + expert demonstration video of task tying, followed by 30 mins	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, 2.04) P = 0.58

			familiarisation with equipment. Practice: 1h daily, 6 days per week for 4 weeks. Performance evaluation: combined assessment of each attempt throughout intervention.											
Olms 2016 Germany	Simulated colour matching teeth	Dental students UGY3	Intervention duration: 1 session Study conducted during 10 wk routine university module on matching tooth shades involving variety of teaching + practice opportunities. Performance evaluation: 2 wks after intervention (within one university module).	1 x 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, 2.73) P < 0.001
Pavo 2016 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 x 2h additional training session: instructional video + training using modified Peyton 4 step approach[c] Brief practice (few mins)	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 & depth, hand position, decompression & hands-off time. Instructional video for intervention group had demonstrated this.		✓	✓			✓	✓				0.25 (-0.02, 0.51) P = 0.06

			in pairs using a manikin. Performance evaluation: end of session.											
Skeff 1983 USA	Clinical teaching skills during ward round	Physicians	Intervention duration:1 session in the middle of 4wk ward duty. At mid & 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 wk later, at end of ward duty	1 x 60 min fb discussion with peer, including video review, trainee ratings & self-assessment to enable physician to identify strengths & devise solutions to problems.	✓				✓	✓	✓		✓	✓	0.56 (-0.15, 1.27) P = 0.12
Soucisse 2017 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 wk later (ongoing clinical work as a surgical resident).	1 x 30min 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, 1.05) P = 0.42

Vafaei 2017 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 x 5min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓					3.04 (1.95, 4.13) P < 0.001
Xeroulis 2007 (fb <u>after</u>) Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	Medical students UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. Performance evaluation: end of session.	Expert fb as needed (expert or learner initiated), <u>after</u> practice trials, involving constructive ways to improve + expert demonstration.	✓			✓		✓	✓				0.86 (-0.08, 1.80) P = 0.06
Xeroulis 2007 (fb <u>during</u>)	Same as above	Same as above	Same as above	Same as above except expert fb <u>during</u> practice trials.	✓		✓			✓	✓				1.44 (0.43, 2.46) P = 0.004

Abbreviations:
% = percentage; CI= confidence interval; CPR = cardiopulmonary resuscitation; GI = gastrointestinal; Info = information; M = male; O&G = obstetrics and gynaecology; PGY = postgraduate year; SMD= standardised mean difference; UGY = undergraduate year; wk =week/s

Footnotes:
a= See 'Meta-analysis' section in Results for additional study details.
b= Machine output information: simulator metrics (e.g. procedural time or instrument path length) or CPR machine information (e.g. compression rate and depth)
[41]c= Peytons' 4 step model[41]

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 - Appendix 2).

[Figures 2 and 3 here]

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.

[Figure 4 here]

Meta-analysis

A meta-analysis of the impact of verbal face-to-face feedback compared to 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 provided *during* practice in one group and *directly after* practice in another were compared to no feedback;[37] in another study, feedback provided by an *expert* in one group and by a *peer* in another group[31] were compared to 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 to no feedback on workplace task performance found a standardised mean difference of 1.09 (95% confidence

interval (CI) 0.59 to 1.59; P < 0. 001) using a random-effects model. The forest plot is presented in Figure 5a.

[Figure 5 here].

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 ten comparisons.[30, 31, 33-37, 39] The standardised mean difference 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.[42] (See 'Certainty of evidence' section in the supplementary material for more details - Appendix 2). Figure 6 displays the Summary of Findings table.

[Figure 6 here]

Comparison 2: The effect of verbal face-to-face feedback, compared to alternative feedback, on performance.

Included studies

Table 2 presents the characteristics of included studies in the comparison of verbal face-to-face feedback compared to alternative feedback. Twenty studies (22 comparisons) were included in this analysis and involved verbal, face-to-face feedback compared to alternative feedback. Nine

studies (9/20, 45%) were published in the last five 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%) males from 13 studies that reported gender data.[33, 38, 43-53] Included studies involved students (medical, mixed health professions and pharmacy) (1869, 95%) in 16 studies,[33, 37, 38, 43-46, 48-56] and doctors (105, 5%) in four studies.[31, 39, 47, 57] All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[39, 47] Three studies evaluated performance in a clinical practice setting (involving teaching skills,[39] professional and communication skills[47] and oral case presentations[56]) 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, 43-46, 48-55, 57] (See 'Included studies' section in the supplementary material for more details - Appendix 3).

Risk of bias

In summarising the risk of bias across domains within each study, two studies were rated as low risk,[44, 51] seven studies were rated as 'high risk',[38, 48, 52, 53, 55, 57] 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 - Appendix 3).

Effect of interventions

Figure 7 presents the forest plot and standardised mean differences (SMD). One additional study[44] 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 odds ratio 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.

[Figure 7 here]

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (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 feedback to an individual health professional	**Intervention B**	**Study outcomes**[a] Unless otherwise stated, effects are SMD (95% CI) P value in favour of feedback intervention A
Al-Jundi[b] **2017** 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 x 10 mins for baseline performance. Performance evaluation: 2 days later	**Immediate face-to-face + written expert feedback** 1 x 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 x 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, -0.79) P < 0.001 favours feedback intervention B
Backstein **2005** Canada	Simulated surgical procedure using a bench top model (vascular anastomosis)	**Doctors in surgical training** PGY1	Intervention duration:4 wk Lecture on surgical procedure.	**Review of performance video with expert fb** **+ practice sessions with expert fb available**	**Practice Sessions with expert fb available**	0.86 (0.05, 1.67) P = 0.03

			3 x 2h 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 wk 4	3 x weekly videotaping of surgical procedure, with expert feedback available during task, followed by up to 15min review of video with expert fb		
Baldwin 2015 England	Simulated BLS	**Health professional students** medical (58%), physio (12%), pharmacy (10%), nursing (10%), dentistry (10%) UGY1 33% M	Intervention duration: 4 wk Instruction and practice with manikin 3 x 2.5h 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 wk 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[c] with final summary.	**'Feedback sandwich model'** Fb involved a point for improvement in between 2 points of praise.	OR 1.25[d] (0.85, 1.84) P = 0.25
Boehler 2006 USA	Simulated surgical skill using a bench top model (tying a 2-handed square knot)	**Medical students** UGY2-3 52% M	Intervention duration: 1 session Instruction in knot tying from surgeon. 1 x baseline performance. Performance evaluation: end of session.	**Expert feedback** 1 x episode of fb from expert surgeon, directly after performance, describing 1-2 specific ways to improve performance.	**Compliment** 1 x pre-scripted general compliment e.g. 'great job!'	0.98 (0.25, 1.71) P = 0.01
Bosse 2015 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 4 step procedural training method[e] 6 x practice trials. Fb 'positively worded', focused on effect of	**High frequency fb** 6 x episodes of fb, given after each practice trial.	**Low frequency practice** 2 x episodes of fb, given after first and last practice trial.	0.81 (0.21, 1.40) P = 0.01

			actions, given directly after performance by senior peer instructors, trained in procedure & fb. Performance evaluation: end of session.			
Boyle 2011 Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty + stent)	**Doctors training in surgery** PGY4+	Intervention duration: 1 session Teaching & expert demonstration. Fb providers had simulator training. 5 x practice trials (each maximum 40min). Performance evaluation: end of session.	**Expert fb** 5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	**Peer fb** 5 x fb episodes. Peer discussed simulator output, any task errors & the teaching instructions given at start of session.	0.46 (-0.70, 1.61) P = 0.41
Brinkman 2007 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 feedback 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 feedback as part of clinical training** 1 x 30min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses & 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 feedback as part of clinical training** Performance ratings from nurses and patients not seen.	2.70 (1.75, 3.64) P < 0.001
DeLucenay 2017 USA	Simulated pharmacist patient consultation (identifying prescription errors	**Pharmacy students** UGY3	Intervention duration: 1 semester. Study conducted during usual university module on medication counselling involving 15 min SP	**Immediate face-to-face fb** 4 x expert fb directly after SP consultation and SP fb, based on expert's direct observation of SP consultation (unseen by participants).	**Delayed written fb** 4 x videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.	0.30 (-0.02, 0.62) P = 0.07

	and communication skills)		consultations, each on a different topic. Directly after each one, SP provided 5min fb on communication skills. Performance evaluation: last 4 SP consultations.	Fb included performance grade, performance and topic discussion with suggested improvements.		
Lee 2016 Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis)	**Medical students** UGY3-4 78% M	Intervention duration: 3 wk Instruction and expert demonstration of procedure, followed by 3 x weekly 30min practice sessions. Performance evaluation: end of 3rd session.	**Early feedback** 1 x 10-15min expert fb directly after first practice attempt, focused on assessment domains.	**Late feedback** Same as early fb but at end of second practice session.	1.3 (0.26, 2.34) P = 0.01
Manzone 2014 (verbal comment focused on performance vs verbal comment + comparison to training levels) Canada	Simulated intubation using manikin	**Medical students** UGY1-2	Intervention duration: 1 session Instructional video on intubation. 1-1.5h practice with manikin, with learner in 4 different positions (5 x practice trials in each position). 10 x fb by expert, given directly after practice trials in 2 positions (2 x 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 2 aspects of performance (either done correctly or not) e.g. 'improper use of the laryngoscope'. + individual's progress on task.	**Performance comment compared to others (different training levels)** Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standard at different training levels e.g. 'your performance was at the level of a resident.'	-0.93 (-1.89, 0.03) P = 0.05
Manzone 2014 (verbal comment on performance vs numerical rating, focused on	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	-0.37 (-1.26, 0.51) P = 0.39

individual progress)					of hand movements). Plotted on graph to focus on own progress	
Manzone 2014 (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 to 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 medical student to specialist	-2.87 (-4.20, -1.55) P < 0.001 favours feedback intervention B
O'Connor 2008 USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk 1st session: 2h instruction and practice suturing & tying knots 'until able to do it easily'. 2nd session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30mins familiarisation with equipment. Practice: 1h 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 & given target goals. .	**Simulator output information**	0.51 (-1.16, 2.19) P = 0.48
Ozcakar 2009	Simulated patient consultation with a simulated patient	**Medical students** UGY2	Intervention duration: 2 wk	**Video review with expert + expert fb**	**Expert fb** 1 x expert fb directly after SP consultation	0.32 (-0.23, 0.87) P = 0.24

Turkey	(communication and history taking skills)	62% M	Study conducted during routine university module on clinical skills training. Evaluation: 2 wk after intervention following clinical skills lectures + practice with video recording.	1 x videotaping of SP consultation. Directly afterwards, review video with expert plus fb.		
Pavo 2016 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 x 2h 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 & depth, hand position, decompression & 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 & depth plus automated audio advice to correct any deviations during CPR.	-0.09 (-0.36, 0.18) P = 0.53
Rogers 2012 USA	Simulated surgical skill (tying a single 2-handed square knot)	**Medical students** 'surgical clerkship year'	Intervention duration: 1 session Training in knot tying. 2 x practice trials (1 before & 1 after training). Performance evaluation: end of session.	**Expert fb** 1x fb from expert, with specific information on improving subsequent performance, directly after performance.	**Compliment** 1 x general compliment from expert, instead of fb.	1.69 (1.06, 2.32) P < 0.001
Skeff 1983 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	**Expert peer fb** 1 x 1h session mid-term 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	**Written fb** Received written summary of trainees' evaluation of physician's teaching skills.	-0.36 (-1.06, 0.34) P = 0.30

			mid- and end of 1 month term.	devise solutions for teaching problems		
Sox 2014 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 x 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 x constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw 1 item evaluation form but not given a copy.	0.15 (-0.07, 0.37) P = 0.17
Strandbygaard 2013 Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extra-uterine pregnancy)	**Medical students** UGY 4-6 44% M	Intervention duration: 2 months 1 x 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 3h) for up to 2 months.	**Standardised expert fb with later, additional expert fb if requested by learner + simulator performance score** 1-3 x 10-12min episodes of expert fb involving information on how to perform task components correctly. 1st fb episode provided after first practice trial; learner could ask for up to 2 additional fb episodes (optional) involving same standardised advice + simulator performance score.	**Simulator performance score**	1.31 (0.86, 1.77) P < 0.001
Van de Ridder 2015a	Simulated hearing test with a simulated patient	**Medical students** UGY1	Intervention duration: 1 x session	**Positively framed fb** 1x fb directly after baseline performance. Fb comment	**Negatively framed fb** 1x episode fb directly after practice trial. Fb comment	0.41 (-0.06, 0.87) P = 0.08

(Advances in Health Science Education) Netherlands	(Weber & Rinne test)	35% M	Instructional video of task. 1 x baseline performance. Fb from senior medical student with acting experience & 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.	started with global praise followed by the most suitable suggestion for improvement, selected from a list of 4 commonest task errors (e.g. 'You did this well; a tip is …')	started with global criticism followed by most appropriate directive advice for improvement, selected from list of 4 commonest task errors. (e.g. 'You did not do this correctly, you should change.')	
Van de Ridder 2015b (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 31% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. All fb providers trained for 1h 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 x fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag & introduced as Professor ENT, wearing a white coat).	**Low credibility fb provider** 1 x fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female, informally dressed)	-0.23 (-0.71, 0.26) P = 0.36
Xeroulis 2007 Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. 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, 0.73) P = 0.96

Abbreviations:
% = percentage; BLS = basic life support; CI= confidence interval; CPR = cardiopulmonary resuscitation; ENT = ear, nose and throat specialist; ERC = European Resuscitation Council; fb= feedback; GRS= global rating scale; h = hour; Max = maximum; min = minutes; NG= nasogastric; NR= not reported; OR = odds ratio; UGY = undergraduate year (referring to university year level); physio= physiotherapy; PGY = postgraduate year (referring to post-qualification year); SMD= standardised mean difference; SP = simulated patients; VR = virtual reality; W&R = Weber & Rinne test; wk = week/s

Footnotes:
a= See Figure 7 forest plot for additional study details;
b = additional data obtained from authors, enabling calculation of mean, SD and % men;
c = Advocacy Inquiry approach;[58]
d = Categorical data only available (see text in Results for more details).
e = Peyton's 4 steps;[41]

DISCUSSION

Comparison 1: The effect of verbal face-to-face feedback, compared to 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 to no feedback, with SMD 0.7 (95% CI 0.37 to 1.03; P < 0.001) 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 - 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.[59] 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 impact 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 reported that 0.5 to 16% more participants followed desired practice when involved in an audit and feedback intervention.[19] In comparison, a meta-analysis by Kluger and DeNisi,[60] which analysed any type of feedback across any discipline, compared to 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 i) effect size 0.55 when feedback included information on any changes since the previous attempt, ii) effect size 0.51 when a specific and challenging goal was set, iii) effect size 0.47 when feedback posed little threat to self-esteem and iv) effect size 0.43 when feedback included information on the correct outcome.[60]

Comparison 2: The effect of verbal face-to-face feedback, compared to alternative feedback, on performance.

For the second comparison of the effect of verbal face-to-face feedback compared to 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 to routine monthly written feedback from supervisors;[47] expert feedback early in a practice period compared to later;[49] additional episodes of feedback from experts;[46, 51] additional episodes of feedback involving expert video analysis[57] and expert feedback compared to compliments.[45, 54]

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, 60] is delivered by educators

with perceived credibility,[18, 19, 61-65] and strategic use of both early and delayed feedback.[49, 61] (See 'Discussion' section in the supplementary material for more details - 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, 41, 49, 61, 66-70] learners' background, task complexity and practice opportunities.[10, 15, 71, 72] (See 'Discussion' section in the supplementary material for more details - 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.[73]

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 to 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 programs 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 to 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,

Declarations of interest: The authors report no declarations of interest.

Author Contributions: CJ and JK designed the review, CJ created the search strategy and undertook the literature searches, CJ and MW undertook study screening, data extraction and

quality assessment with oversight from JK, CJ and JK undertook data analysis and interpreted the findings, CJ prepared the manuscript and all authors contributed to the final version.

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

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: The authors report no competing interests.

Patient consent: Not required.

Ethics Approval: Not required.

Data sharing statement: Data access can be requested by contacting the corresponding author.

Figure and Table Legends

Figure 1: PRISMA flow diagram.

Figure 2: Risk of bias graph

Figure 3: Risk of bias summary

Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 6: Summary of findings table for the effect of verbal face-to-face feedback, compared to no feedback, on performance.

Figure 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance.

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

Table 2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

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. *Academic Medicine* 2016;91(5):645-9. doi: 10.1097/acm.0000000000001048 [published Online First: 2015/12/18]
3. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34(3):205-14. doi: 10.3109/0142159x.2012.652239 [published Online First: 2012/03/01]
4. Holmboe ES. Realizing the promise of competency-based medical education. *Academic Medicine* 2015;90(4):411-3. doi: 10.1097/acm.0000000000000515 [published Online First: 2014/10/09]
5. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association* 2009;302(12):1316-26. doi: https://dx.doi.org/10.1001/jama.2009.1365
6. Archer JC. State of the science in health professional education: effective feedback. *Medical Education* 2010;44:101-08. doi: 10.1111/j.1365-2923.2009.03546.x
7. Bing-You RG, Hayes V, Varaklis K, et al. Feedback for learners in medical education: what is known? A scoping review. *Academic Medicine* 2017;92(9):1346-54. doi: https://dx.doi.org/10.1097/ACM.0000000000001578
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 Medical Education* 2016;16(1):96. doi: 10.1186/s12909-016-0613-5
9. Watling C, Ginsburg S. Assessment, feedback and the alchemy of learning. *Medical Education* 2019;53(1):76-85. doi: 10.1111/medu.13645 [published Online First: 2018/08/04]
10. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist* 2002;57(9):705-17. [published Online First: 2002/09/20]
11. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine* 2015;90(5):609-14. doi: https://dx.doi.org/10.1097/ACM.0000000000000560

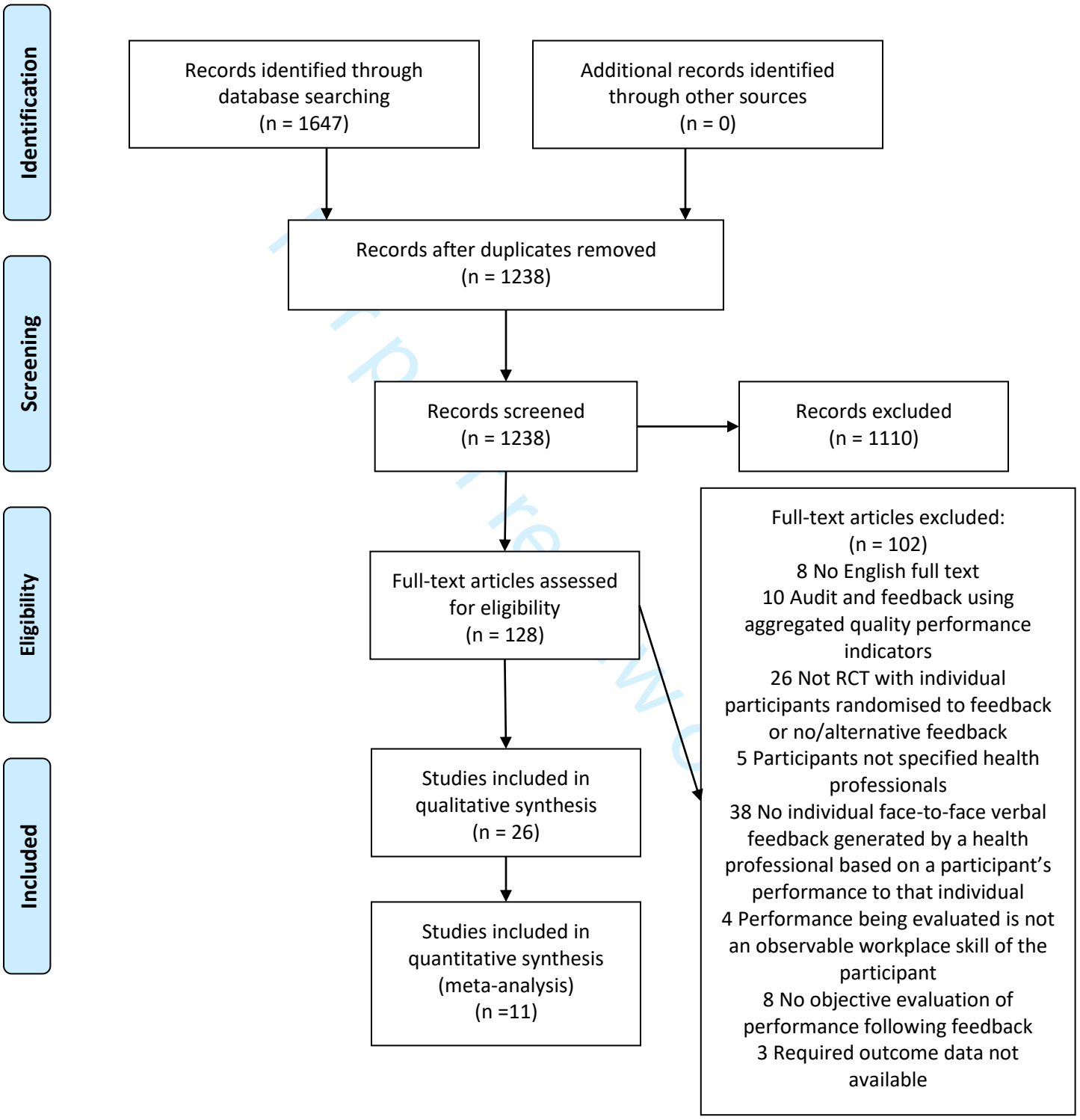
12. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Medical Teacher* 2011;33(12):961-73. doi: https://dx.doi.org/10.3109/0142159X.2011.595435
13. Johnson CE, Keating JL, Farlie MK, et al. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. *BMC Medical Education* 2019;19(1):129. doi: 10.1186/s12909-019-1524-z
14. Hattie J, Timperley H. The Power of Feedback. *Review of Educational Research* 2007;77(1):81-112. doi: 10.3102/003465430298487
15. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
16. Molloy E, Boud D. Changing conceptions of feedback. In: D. B, Molloy E, eds. Feedback in higher and professional education. London: Routledge 2013:11-33.
17. Ende J. Feedback in clinical medical education. *Journal of the American Medical Association* 1983;250(6):777-81. doi: 10.1001/jama.1983.03340060055026
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. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
19. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
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(4):264-9, w64. doi: 10.7326/0003-4819-151-4-200908180-00135 [published Online First: 2009/07/23]
21. Higgins JPT, Churchill R, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). Available from www.training.cochrane.org/handbook: Cochrane, 2017.
22. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *British Medical Journal* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
23. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List. *Journal of Clinical Epidemiology* 1998;51(12):1235-41. doi: 10.1016/S0895-4356(98)00131-0
24. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal* 2011;343 doi: 10.1136/bmj.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). Available from www.training.cochrane.org/handbook: Cochrane, 2017.
26. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal* 2011;342:d549.
27. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration 2011 [Available from: www.handbook.cochrane.org.
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). Available from www.training.cochrane.org/handbook: Cochrane, 2017.
29. GRADE guidelines Journal of Clinical Epidemiology series [Available from: https://www.gradeworkinggroup.org/2020.

30. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
31. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
32. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
34. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
35. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
36. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
37. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
38. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
39. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
40. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
41. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
42. 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. *Journal of Clinical Epidemiology* 2013;66(2):151-57. doi: 10.1016/j.jclinepi.2012.01.006
43. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
44. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
45. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
46. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5

47. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine* 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44
48. 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(1):322-28.
49. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
50. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
51. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e
52. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
53. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
54. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009
55. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397
56. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209
57. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
58. 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 Medical Simulation* 2016;1(1):49-55.
59. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates 1988.
60. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
61. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
62. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
63. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
64. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.

65. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
66. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.
67. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360
68. Elliott SN. Educational psychology: effective teaching, effective learning. 3rd ed. Boston: McGraw-Hill 2000.
69. Wadsworth BJ. Piaget's theory of cognitive and affective development: Foundations of constructivism. 5th ed. White Plains: Longman Publishing 1996.
70. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
71. 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 ed. Oxford: Wiley Blackwell 2014:7-29.
72. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227-68. doi: 10.1207/s15327965pli1104_01
73. Belfield C, Thomas H, Bullock A, et al. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;23(2):164-70. doi: 10.1080/0142150020031084 [published Online First: 2001/05/24]

Figure 1: PRISMA flow diagram for systematic review of verbal face-to-face feedback compared to no or alternative feedback.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

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

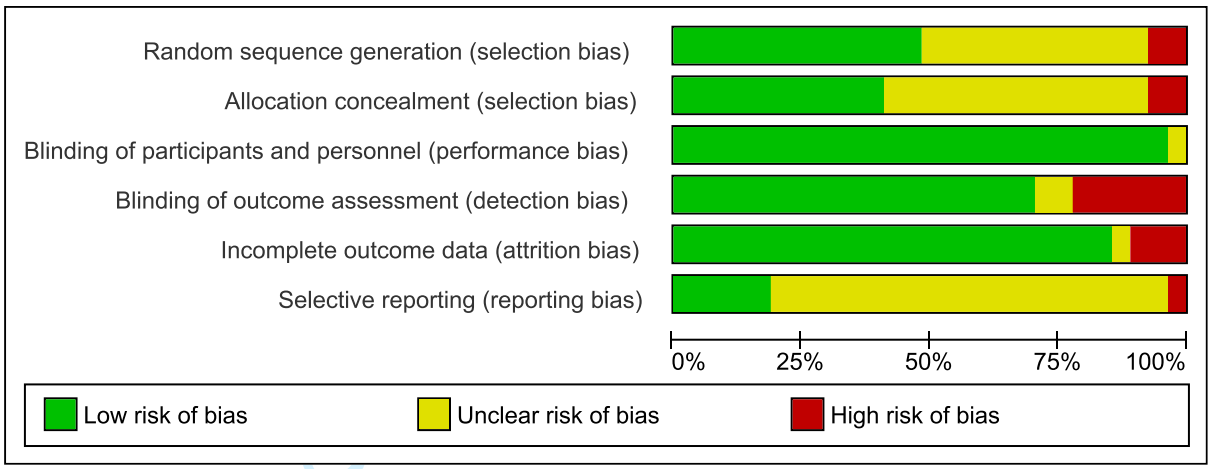


Figure 3: Risk of bias summary: review authors' judgements about each risk of bias item for each included study

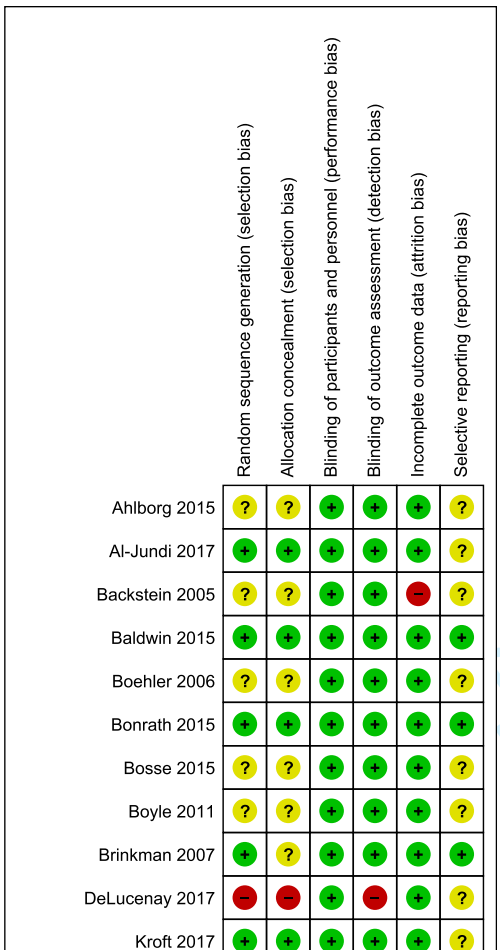
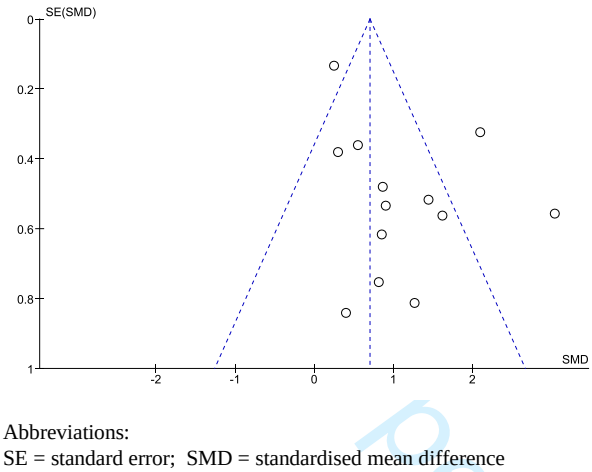
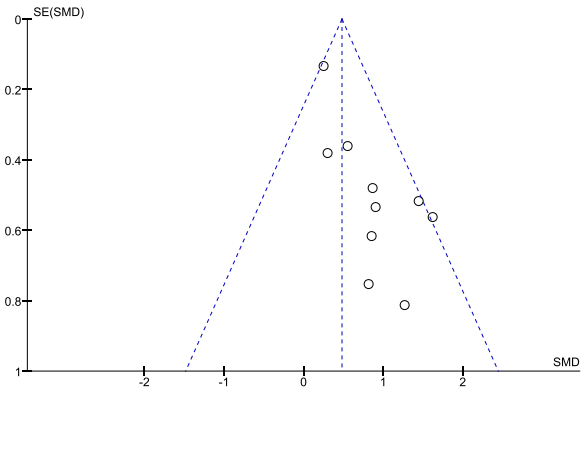


Figure 4: Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

a) all included studies



b) excluding studies at high risk of bias



Abbreviations:
SE = standard error; SMD = standardised mean difference

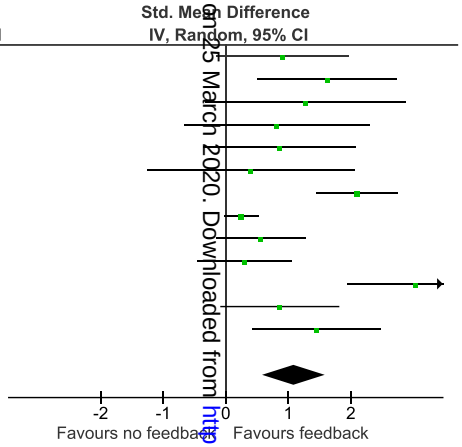
Footnote:
Meta-analysis calcuated using a fixed effects model.
The dotted vertical line represents the overall effect estimate and the dotted slanted lines represent the 95% confidence interval.

Figure 5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance

a) All included studies

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.6%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.3%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	5.3%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	5.7%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.9%	0.85 [-0.35, 2.06]
O'Connor 2008	-1.96	1.25	3	-2.5	0.9	3	5.1%	0.40 [-1.25, 2.04]
Olms 2016	-1.1	1.8	30	-7.4	3.8	30	9.6%	2.09 [1.45, 2.73]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	11.0%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	9.2%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	9.1%	0.30 [-0.44, 1.05]
Vafei 2017	17.73	1.09	15	12.13	2.29	15	7.4%	3.04 [1.95, 4.13]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	8.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	7.7%	1.44 [0.43, 2.46]
Total (95% CI)			**253**			**235**	**100.0%**	**1.09 [0.59, 1.59]**

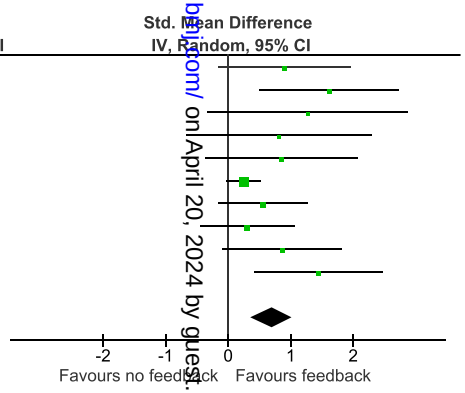
Heterogeneity: Tau² = 0.58; Chi² = 54.26, df = 12 (P < 0.00001); I² = 78%
Test for overall effect: Z = 4.26 (P < 0.0001)



b) Excluding studies at high risk of bias (sensitivity analysis)

Study or Subgroup	Feedback Mean	SD	Total	No feedback Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ahlborg 2015	-3.9	1.85	8	-5.9	2.3	8	7.7%	0.91 [-0.14, 1.95]
Bonrath 2015	4.5	1	9	2.8	1	9	7.1%	1.62 [0.52, 2.72]
Boyle 2011 (expert fb)	-16	10	6	-38	24	3	3.8%	1.27 [-0.32, 2.87]
Boyle 2011 (peer fb)	-22	14	6	-38	24	3	4.4%	0.81 [-0.66, 2.29]
Kroft 2017	28.92	8.86	6	19.67	11.07	6	6.2%	0.85 [-0.35, 2.06]
Pavo 2016	0.33	0.27	110	0.27	0.21	114	27.9%	0.25 [-0.02, 0.51]
Skeff 1983	3.37	0.5	16	3.09	0.48	16	13.3%	0.56 [-0.15, 1.27]
Soucisse 2017	22.57	4.83	14	20.93	5.65	14	12.4%	0.30 [-0.44, 1.05]
Xeroulis 2007 (fb after)	16.9	5	15	12.71	3.77	7	9.1%	0.86 [-0.08, 1.80]
Xeroulis 2007 (fb during)	16.97	2.33	15	12.71	3.77	7	8.1%	1.44 [0.43, 2.46]
Total (95% CI)			**205**			**187**	**100.0%**	**0.70 [0.37, 1.03]**

Heterogeneity: Tau² = 0.08; Chi² = 13.64, df = 9 (P = 0.14); I² = 34%
Test for overall effect: Z = 4.14 (P < 0.0001)



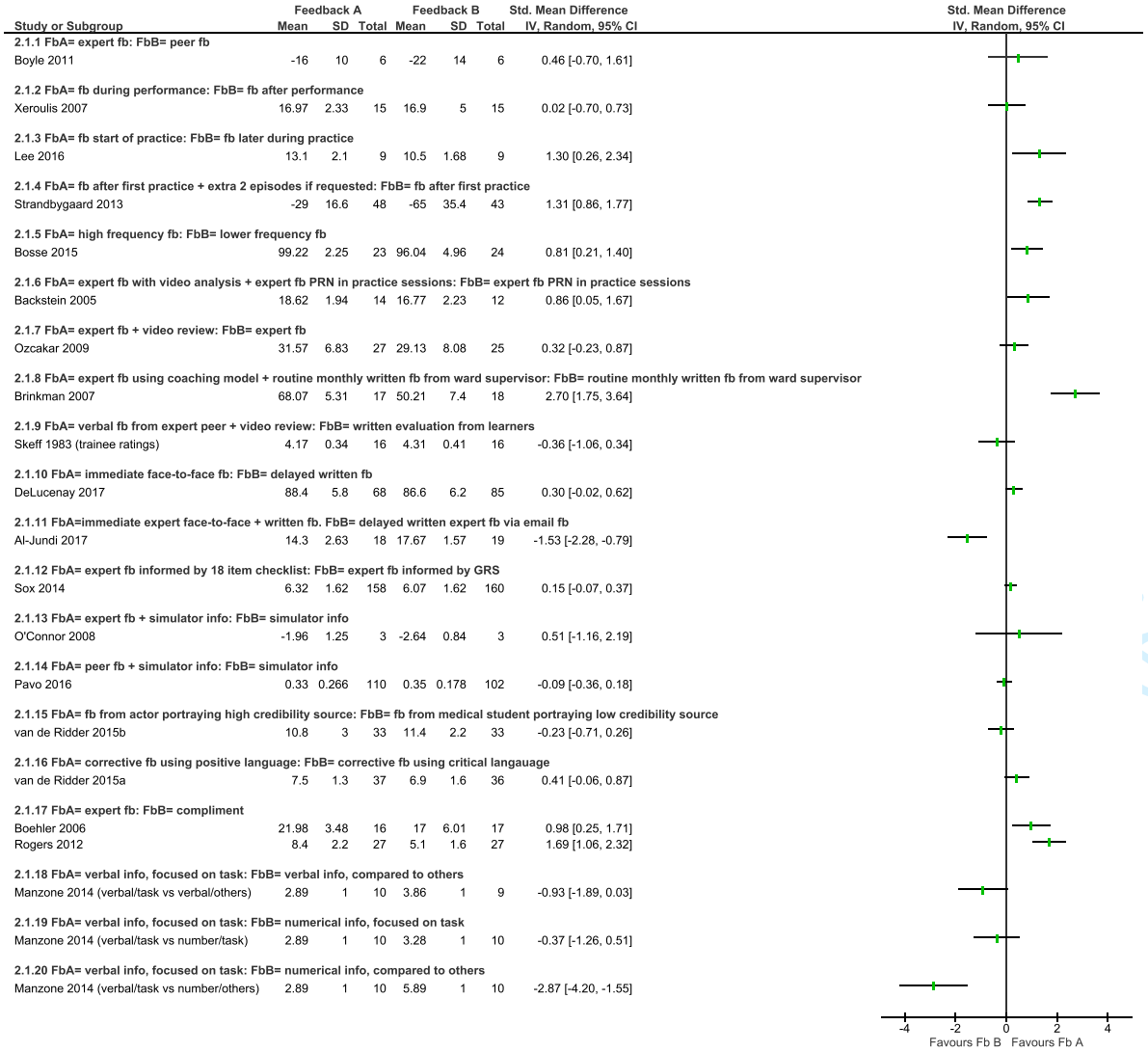
Abbreviations:
SD = standard deviation; CI = confidence interval; SMD = standardised mean difference

Footnotes:
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

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

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 7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance



Abbreviations:
Fb= feedback; GRS= global rating scale; info= information; PRN= 'as required'; SD = standard deviation; CI=confidence interval

Footnotes:
[1]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

http://bmjopen-2019-030672 on 25 March 2020. Downloaded from

Database(s): Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:

#	Searches	Results
1	*Feedback/	6031
2	Feedback, Psychological/	3311
3	Formative Feedback/	467
4	(feedback adj3 (effective or formative or constructive or quality or clinical or performance)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4860
5	1 or 2 or 3 or 4	13233
6	exp Health Personnel/	470058
7	exp Health Occupations/	1648689
8	exp Dentistry/	386159
9	exp Social Work/	17331
10	exp Psychology/	66579
11	Occupational Therapy/	13213
12	Radiotherapy/	42757
13	Radiography/	334082
14	Mentors/	9949
15	exp Students, Health Occupations/	60760
16	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	2722708
17	clinician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	171551
18	(health* adj2 (staff or personnel or faculty or provider* or worker* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	303659
19	doctor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104240
20	physician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	495037
21	(medical adj3 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	133207
22	general practitioner*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45015
23	(general pract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	45488
24	(family adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21549
25	(primary care adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10661
26	(primary health* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2156
27	(registrar or registrars or senior house officer* or resident or residents or hospital medical officer* or intern or interns or house officer*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	129452
28	dentist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	115221
29	(dent* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21015
30	nurs*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	629304
31	(midwife or midwives).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	18155

32	(midwife* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	965
33	(allied health adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	12936
34	physiotherapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5348
35	physical therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5123
36	(physiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	731
37	(physical therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	647
38	occupational therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4500
39	(occupational therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1058
40	speech therap*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7630
41	(speech therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	22
42	speech language therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	136
43	(speech language therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	14
44	speech pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	656
45	(speech pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	31
46	speech language pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1561
47	(speech language pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	105
48	dietician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1352
49	(dietetic* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	603
50	podiatrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	626
51	(podiatry adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	19
52	chiropodist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	119
53	(chiropody adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2
54	social worker*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8079
55	(social work* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8517

56	psychologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	11986
57	(psychology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1532
58	(osteopath* or osteopathic physician*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	6480
59	(osteopath* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	296
60	chiropractor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1148
61	(chiropract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	241
62	pharmacist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	27316
63	(pharmac* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7184
64	optometrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1764
65	(optometr* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	290
66	(Radiographer* or radiological technologist* or radiation therapist* or radiotherapist* or radiation therapy technologist*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2449
67	(radiograph* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	267
68	(radiation therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	53
69	(radiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	100
70	(supervisor* or tutor* or trainer* or educator* or teacher* or mentor* or preceptor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	104110
71	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70	1846616
72	16 or 71	3625236
73	exp Education, Professional/	284462
74	exp Educational Measurement/	137138
75	exp Professional Practice/	247602
76	exp Simulation Training/	6239
77	(effect* or evaluat* or outcome* or assess* or measur*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10754772
78	73 or 74 or 75 or 76 or 77	11134144
79	randomized controlled trial.pt.	505181
80	controlled clinical trial.pt.	100406
81	randomized.ab.	391590
82	randomly.ab.	266043
83	systematic review.ab,ti.	85419
84	79 or 80 or 81 or 82 or 83	966230
85	5 and 72 and 78 and 84	821
86	limit 85 to (english language and humans)	809

Supplementary material. Appendix 2

RESULTS: SUPPLEMENTARY INFORMATION

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

Included studies

Participants

Participants included 290 (60%) medical students in four studies,[1-4] 60 (12%) dental students in one study[5] and 138 (28%) doctors (doctors training in surgery in three studies,[6-8] training in obstetrics and gynaecology in one study[9] and training in emergency medicine in one study,[10] and physicians in one study[11]).

Participants were novices to the assessed task in five studies (5/11, 45%);[1, 2, 4, 5, 7] and had prior experience in six studies.[3, 6, 8-11]

Workplace tasks and Settings

All studies evaluated performance of a discrete task; there were no longitudinal evaluations. The task occurred in simulation settings in seven studies (7/11, 64%) and clinical practice in four studies (4/11, 36%). The task was a surgical procedure in seven studies (7/11, 64%). Five studies involved simulated surgical tasks including bench top models for knot tying[4] and forming a bowel anastomosis;[8] using a laparoscopic simulator for suturing and knot tying;[2] and using a virtual reality (VR) simulator for laparoscopic surgery[1] and endovascular surgery.[7] Two studies involved laparoscopic surgery in clinical practice.[6, 9] The remaining four studies evaluated simulated matching of tooth colour in a dental school,[5] simulated cardiopulmonary resuscitation (CPR),[3] chest ultrasound for emergency trauma patients[10] and teaching skills in clinical practice.[11]

Feedback Interventions

The feedback source involved a subject expert in all comparisons except two, including one that compared peer feedback with no feedback,[3] and one that compared expert feedback, peer feedback and no feedback.[7] Feedback occurred while the participant performed the task (during) in one study,[3] both during and directly afterwards in two studies,[1, 2] directly afterwards in four studies,[5, 7, 9, 10] after a delay in three studies[6, 8, 11] and one study compared feedback during, feedback directly afterwards and no feedback.[4] In addition to evaluative performance information (as per inclusion criteria), the feedback included corrective advice in all studies except one[10] and one where it was unclear.[7] Feedback included additional information from a simulator in three studies,[1, 2, 7] a video of the participant's performance in two studies[6, 11] and written performance information in two studies.[5, 11]

Teaching and Practice

In addition, instruction and expert demonstration of the task were provided in six studies (6/11, 55%), including all five studies involving novice participants[1, 2, 4, 5, 7] and one study that involved CPR for medical students, many of whom had previously attended a course.[3] The other five studies involved doctors working in clinical practice; in these studies, no instruction or expert demonstration was included within the research intervention but may or may not have occurred during the course of routine work during that time. One study involved physicians' teaching on ward rounds[11] and the other four studies assessed tasks by doctors training in relevant specialties.[6, 8-10]

The amount of practice varied substantially between different studies, for both simple and complex tasks. For example, comparing two studies that involved simple surgical knot tying: in Xeroulis,[4] participants had 18 practice attempts in one session and in O'Connor,[2] they could practice up to an hour a day, for 24 days. Looking at more complex surgical procedures, such as simulated surgery using a virtual reality (VR) simulator: in Ahlborg,[1] participants had two

practice attempts at the simulated surgery (laparoscopic salpingectomy) and in Boyle,[7] participants had five attempts at the simulated surgery (renal artery angioplasty and stenting) before the performance evaluation.

Intervention period

The intervention period ranged from one day (most common) up to two months.[6] Nine (9/11, 82%) studies involved a single session (involving one episode of feedback in five studies[5, 8-11] and multiple episodes of feedback in four studies[1, 3, 4, 7]). Two studies (2/11, 18%) had a longer intervention period involving multiple feedback sessions: one study[6] included approximately four coaching sessions regarding bariatric surgery across a two month surgical attachment, and another[2] included almost daily one hour practice sessions for laparoscopic suturing, with feedback throughout each one, over four weeks.

The timing of the post-feedback performance assessment, in relation to the intervention, differed. It occurred directly following the intervention in seven studies: at the end of the single session in five studies[1, 3, 4, 7, 9] and at the end of an extended intervention period in two studies.[2, 6] In the other four studies, the post-feedback performance assessment occurred some weeks after the intervention was completed but while relevant exposure to possible teaching and/or practice opportunities continued. Olms[5] included a single feedback session, with the final evaluation two weeks later, in the midst of a routine one month university teaching unit on tooth shade matching. Skeff[11] arranged a single coaching session on ward round teaching in the middle of physicians' four week ward duty, with the final evaluation post-performance evaluation at the end. Soucisse[8] also organised a single coaching session for surgical residents, with the final evaluation occurring three weeks later. Vafaei[10] involved a single workplace-based assessment with feedback for doctors training in emergency medicine on chest ultrasound for emergency trauma patients, followed by a two month period of routine clinical work before the post-feedback assessment.

Research funding

Regarding research funding, one study[3] that focused on cardiopulmonary resuscitation (CPR) quality, was loaned a device (used to measure CPR parameters and provide automated feedback to participants) for the period of the study by Philips but the company was not otherwise involved in the research; five studies received funding from independent institutions,[1, 4, 6, 9, 11] three studies did not receive any funding[5, 7, 10] and two studies did not report information on funding.[2, 8]

Risk of bias

Five trials described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[3, 5, 6, 8, 9] The other six trials simply stated participants were 'randomised' and had no information on allocation concealment, so we rated these studies as 'unclear'. We analysed baseline performance because, although randomisation removes the need to check comparability in baseline task performance for intervention and comparison groups, it may be useful to check this when participant numbers are small and performance improvement is more likely when baseline performance is low.[12] Seven studies reported no statistically significant differences between baseline performances for the comparison groups.[4, 5, 8-11] and four studies did not report baseline task performance.[1-3, 7] The participants and research team members were not blinded in any included studies because the intervention involved feedback between a research team member and a participant, consistent with most education interventions. However, in all included studies, we thought this was not likely to influence the outcome (post-intervention performance assessment) because implementation and adherence to the intervention were not affected. In eight studies the outcome was assessed by either blinded assessors who rated videos of the participants' performance[4, 6-9, 11] or by a machine (simulator or CPR machine),[1, 3] so we rated these as 'low risk' of bias. In three studies, the feedback provider and outcome assessor appeared to be the

same person, so these were rated as 'high risk'.[2, 5, 10] Across all the studies, the follow up rate for each group was at least 85%. Only two studies had a prior published protocol in addition to reporting all outcomes as planned.[6, 8] For all other studies, it could not be ascertained if outcomes had been selectively reported, so these were rated as 'unclear', except one. This one study was rated as 'high risk' for selective outcome reporting because it did not include the expected information on performance post-intervention.[2]

In summarising the risk of bias across domains within each study, two studies had all domains rated 'low risk, so these were rated low risk.[6, 8] Six studies had at least one domain with 'unclear' risk but no 'high risk' ratings, so these were rated as 'unclear' risk of bias.[1, 3, 4, 7, 9, 11] Three studies had at least one domain at high risk of bias, so we judged these studies to be at 'high risk' of bias.[2, 5, 10]

Certainty of evidence

For the comparison of verbal face-to-face feedback compared to no feedback, excluding studies at high risk of bias, we graded the quality of evidence for the outcome of 'objective assessment of a health professional's performance'. The risk of bias was rated as 'unclear' across multiple included studies and the overall body of evidence indicated this was likely to seriously alter the results, so we downgraded the overall evidence by one level. The two aspects that were most influential on our decision were the lack of allocation concealment and prior published protocols to preclude selective reporting of outcomes. Participant and research team member blinding was not possible due to the intervention. However, this had limited impact on the selected outcome 'objective assessment of performance', as no changes occurred in intervention implementation or adherence as a consequence of this lack of blinding.[13] We judged the results to be directly applicable to our review question and therefore the evidence was not downgraded for indirectness. There was some methodological and statistical heterogeneity across studies (the test for heterogeneity was not significant with $P = 0.14$ and I^2

= 34%), which was not explained by subgroup analysis. However, all studies reported a beneficial effect, so the uncertainty seemed to lay in the magnitude of effect rather than the presence of an effect. Therefore, we decided not to downgrade the evidence due to inconsistency.[14] We judged the effect size to be sufficiently precise and therefore did not downgrade the evidence for imprecision of results. This was based on sufficient numbers of participants (392 when studies with high risk of bias were excluded) and a consistent beneficial effect, indicated by the confidence interval for the overall effect estimate not crossing zero and all individual studies showing a beneficial effect with substantial overlap in their confidence intervals. Finally, we judged that there was likely to be a systematic overestimation of the underlying beneficial effect of feedback because we strongly suspected publication bias (see Funnel plot 5b) and therefore we downgraded the evidence by one level.

In summary, combining all five GRADE criteria for assessing the certainty of evidence, we downgraded the overall rating by one, from high to low. We judged that the quality of the evidence was low contributing to the effect estimate of 0.70 in the comparison of verbal face-to-face feedback to no feedback after excluding studies with a high risk of bias. Hence face-to-face feedback may result in a moderate to large improvement in health professionals' workplace task performance.

References

1. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education* 2015;6:93-100. doi: https://dx.doi.org/10.5116/ijme.55a2.218b
2. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
3. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
4. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012

5. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry* 2016;28(5):277-86. doi: 10.1111/jerd.12231
6. Bonrath EM, Dedy NJ, Gordon LE, et al. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery* 2015;262(2):205-12. doi: https://dx.doi.org/10.1097/SLA.0000000000001214
7. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
8. Soucisse ML, Boulva K, Sideris L, et al. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education* 2017;74(2):365-71. doi: https://dx.doi.org/10.1016/j.jsurg.2016.09.002
9. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education* 2017;9(2):190-94. doi: https://dx.doi.org/10.4300/JGME-D-16-00238.1
10. Vafaei A, Heidari K, Hosseini MA, et al. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency* 2017;5 (1) (no pagination)(e28)
11. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.
12. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
13. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *British Medical Journal* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]
14. 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. *Journal of Clinical Epidemiology* 2013;66(2):151-57. doi: 10.1016/j.jclinepi.2012.01.006

Supplementary material. Appendix 3

Results: SUPPLEMENTARY INFORMATION

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

Included studies

Participants

Included studies involved medical students (1076, 55%) in 14 studies,[1-14] mixed health professional students (640, 32%) in one study,[15] pharmacy students (153, 8%) in one study[16] and doctors (105, 5%) in four studies.[17-20]

Participants were novice to the task in 11 studies (11/20, 55%). Three studies documented prior experience: one study involved attending physicians teaching on ward rounds with a range of experience[20] and two studies documented previous training including CPR[12] and history taking and communication skills in medical students.[11] The remaining six studies did not report this information. One of these studies[15] involved teaching CPR to first year health professional students across a mix of disciplines, some of whom may have had prior experience. One study[19] involved evaluating professional and clinical skills in first year paediatric residents who likely had relevant training as medical students. In two of these studies, the baseline performance of junior medical students' attempting surgical knot tying was poor, which suggest limited prior experience.[2, 3] In the last two studies there was no information on prior experience: one assessed a simulated medication consultation by third year pharmacy students[16] and another[9] assessed case presentation skills in third year medical students in their paediatric attachment.

Workplace tasks and Settings

All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[19, 20] Three studies evaluated performance in a clinical practice setting

(involving teaching skills[20] professional and communication skills[19] and oral case presentations[9]) and the remaining 17 assessed performance in a simulated environment.[1-8, 10-18] Simulated surgical tasks included suturing and/or knot tying,[1-3, 10, 14] bench top surgical procedures such as vascular anastomosis,[18] flexible ureteroscopy for urolithiasis,[5] renal artery angioplasty and stent placment,[17] or surgery using a VR simulator for a laparoscopic salpingectomy.[13] Simulated critical care tasks included basic life support (BLS)/CPR,[12, 15] intubation[6] and pharmacist-patient consultation.[16] The remaining simulated tasks included a hearing test,[7, 8] simulated patient consultation[11] and nasogastric tube insertion.[4]

Interventions

Each study included at least one verbal face-to-face feedback group, in accordance with the inclusion criteria.

Some studies investigated straightforward variations in feedback, including differences in frequency (low or high[4]), stage of practice (early or late[5]), different feedback models ('learning conversation' compared with 'feedback sandwich' frameworks[15]), source expertise (expert or peer[17]) and expert feedback compared to compliments.[2, 3] Another collection of studies explored the effect of adding expert feedback to other interventions, such as in addition to simulator performance data[12, 13] or to written feedback;[19, 20] or adding expert review of a participant's performance video to a practice session in which expert feedback was available.[18] One study[1] compared verbal feedback by an expert who had just directly observed the performance, with written feedback emailed later that day by another expert who watched a video of the performance. Other studies explored more complex phenomena. One study[6] compared two feedback variations in different combinations across four groups. One variation compared an evaluative verbal comment from an expert, to a written numerical performance rating. The second variation involved an individual comparing their performance evaluation to either their own previous attempts (highlighting individual progress) or to expected

performance at student, resident or specialist level (comparison with others). Another research group investigated two complex influences in separate studies. One study[8] examined how the credibility of the feedback provider (high or low credibility) influenced learner outcomes. The other study[7] examined the effect of phrasing corrective information in different ways, so in one intervention corrective information was framed within a positive phrase whereas in the other, it was framed within a critical phrase.

Research funding

One study[12] was loaned a device by Philips as detailed earlier, seven studies received funding from independent institutions,[4, 6, 9, 13, 14, 19, 20] six studies did not receive any funding[2, 5, 7, 8, 11, 17] and six studies did not report information on funding.[1, 3, 10, 15, 16, 18]

Risk of bias

The risk of bias assessment for the comparison of verbal face-to-face feedback to alternative feedback is presented in Figure 3. Seven described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[1, 7-9, 12, 13, 15] Two studies had adequate random sequence generation, which we rated 'low risk' but had insufficient information on allocation concealment, which we rated 'unclear risk'.[11, 19] The remaining studies simply stated participants were 'randomised' and had insufficient information on allocation concealment, both of which we rated 'unclear risk'. Two studies described inconsistencies with randomisation, so these were rated 'high risk' of bias for sequence generation and allocation concealment.[6, 16] There was unequal baseline performance between groups reported in one study[8] and identified from another study's data (obtained from authors).[1] No statistically significant differences in baseline performance between groups were reported in seven studies [2, 4, 5, 8, 14, 19, 20] and baseline performance was not reported in eleven

studies.[3, 6, 9-13, 15-18] None could blind participants or research team members due to the face-to-face feedback interventions. However we thought this was not likely to influence the outcome as implementation and adherence to the intervention were not affected in all studies, which were rated 'low risk', except one in which some participants may not have experienced the intervention they were allocated to, so it was rated 'unclear'.[9] The outcome was assessed by blinded assessors or machines in all studies, which were rated 'low risk' except two studies that did not explicitly describe blinded assessors, which were rated 'unclear'[8, 20] and four studies that seemed to have assessors who were aware of participant allocation, so these were rated 'high risk'.[10, 11, 16] All had high proportions of participant completion data except three[7, 8, 18] and one report provided insufficient information.[3] Three studies had prior published protocols and reported all outcomes as planned, so they were rated 'low risk'.[13, 15, 19] All of the others did not have a prior published protocol but did present outcomes as expected and were rated as 'unclear',[1-9, 11, 12, 14, 16-18, 20] except one study which was rated as 'high risk'.[10]

In summarising the risk of bias across domains within each study, two studies were rated as low risk[13, 15] as all domains were rated as 'low risk' of bias, seven studies were rated as 'high risk' because at least one domain was rated as 'high risk',[6-8, 10, 16, 18] and the remaining studies were rated as 'unclear' as they had at least one domain with 'unclear' risk but no 'high risk' ratings.

References

1. 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. *Journal of Surgical Education* 2017;74(3):390-97. doi: https://dx.doi.org/10.1016/j.jsurg.2016.11.011
2. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education* 2006;40(8):746-9. doi: https://dx.doi.org/10.1111/j.1365-2929.2006.02503.x
3. Rogers DA, Boehler ML, Schwind CJ, et al. Engaging medical students in the feedback process. *American Journal of Surgery* 2012;203(1):21-5. doi: https://dx.doi.org/10.1016/j.amjsurg.2011.07.009

4. 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 Medical Education* 2015;15:22. doi: https://dx.doi.org/10.1186/s12909-015-0286-5
5. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
6. Manzone J, Tremblay L, You-Ten KE, et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education* 2014;48(4):430-40. doi: https://dx.doi.org/10.1111/medu.12397
7. Ridder J, Peters C, Stokking K, et al. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education* 2015;20(3):803-16. doi: 10.1007/s10459-014-9567-8
8. van de Ridder J, Berk FC, Stokking KM, et al. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher* 2015;37(8):767-74. doi: http://dx.doi.org/10.3109/0142159X.2014.970617
9. Sox CM, Dell M, Phillipi CA, et al. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics* 2014;134(5):965-71. doi: https://dx.doi.org/10.1542/peds.2014-1209
10. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy* 2008;22(7):1614-9. doi: https://dx.doi.org/10.1007/s00464-007-9645-6
11. 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. doi: https://dx.doi.org/10.1186/1471-2458-9-474
12. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
13. 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. *Annals of Surgery* 2013;257(5):839-44. doi: https://dx.doi.org/10.1097/SLA.0b013e31827eee6e
14. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
15. Baldwin LJ, 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. doi: https://dx.doi.org/10.1016/j.resuscitation.2015.08.010
16. 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(1):322-28.
17. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058
18. Backstein D, Agnidis Z, Sahdu R, et al. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery* 2005;48(3):195-200.
19. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of*

Pediatrics & Adolescent Medicine 2007;161(1):44-9. doi: https://dx.doi.org/10.1001/archpedi.161.1.44

20. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine* 1983;75(3):465-70.

Supplementary material. Appendix 4

DISCUSSION

Comparison 1: The effect of verbal face-to-face feedback, compared to no feedback, on performance: supplementary information

Included studies involved health professional students and clinicians (mainly medical) performing a range of workplace tasks, particularly surgical and most commonly in a simulated environment. The meta-analysis results are dominated by one study,[1] evaluating effective compressions during CPR, which contributed the largest number of participants from a single study to the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02, 0.51). Several factors may have contributed to the relatively small overall benefit from this feedback intervention compared to many of the other included studies. These include a short practice period with feedback from a peer (as opposed to an expert) who was concurrently performing a different task (the student performing ventilation provided advice on correcting compressions to the student performing compressions).

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

Previous research has identified that feedback is more effective when the content includes information that makes the goal clear (for example, describing correct performance or providing an expert demonstration of the task) and advice on how to improve.[2-4] However, detailed specifications about feedback content were often not clearly reported in included studies, which suggests that researchers may not have realised the importance of this. The feedback source was more often experts than peers, in our included studies. One small study[5] directly compared expert feedback to peer feedback for novices learning a surgical task using a visual reality simulator. It did not find a statistically significant difference (SMD 0.46, 95%

CI -0.7, 1.61), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother. In earlier systematic reviews[2, 6] and other research,[7-11] feedback from a highly credible source (expert feedback) has been reported to be more effective. Also, the timing of feedback in included studies varied; it was provided while the learner undertook the task, immediately afterwards or some time afterwards. One small study,[12] in which novices learnt to suture, feedback during the task was compared to feedback immediately after each attempt. It did not find a statistically significant difference in performance after one hour of practice but did a month later (beyond the scope of our review), in favour of feedback immediately after practice. In another study, in which students practised simulated laparoscopic surgery, the effect of additional expert feedback was compared with performance information provided by the simulator alone. The authors reported that more participants in the 'simulator feedback only' stopped practising. Previous research has noted that for novices learning a complex task, early feedback and assistance may prevent extreme frustration and giving up.[7] Feedback during task performance results in faster initial skill acquisition compared to feedback after task performance, particularly for procedural skills, as errors are corrected in real-time, but poorer subsequent independent performance.[7, 13] It is thought that a learner develops a mental schema depicting how to do the task, which they develop during practice attempts and this is utilised for subsequent performances.[14, 15] However feedback during task performance appears to interfere with this process, possibly due to cognitive overload.[16, 17]

In the second analysis, verbal face-to-face feedback was compared to a multitude of alternative feedback interventions. In addition to feedback source, frequency, timing and content, there were differences across feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or 'feedback sandwich'), phrasing of feedback (expressing the same corrective information in a positively or negatively

couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into components that might influence feedback outcomes.

Influences on performance due to factors beyond feedback

Firstly, teaching and expert demonstration were common (but not standard) and the amount and type varied across studies, which have previously been shown to impact performance.[18, 19] Practice opportunities also differed enormously across included studies, even those involving similar tasks. In addition, there was variation across learners' prior relevant expertise (e.g. first year medical students or surgical trainees learning a surgical task) and the complexity of the task (knot tying or laparoscopic bariatric surgery). Previous research has shown that teaching and expert demonstration assist a learner to improve, particularly in the initial phases of skill acquisition[18, 19] and practise is essential for mastering any skill.[20, 21] Furthermore, learners who are motivated to learn a challenging but achievable skill are most likely to improve their performance, according to 'goal setting' and 'self-determination' theories.[4, 22]

REFERENCES

1. 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. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine* 2016;24:70. doi: https://dx.doi.org/10.1186/s13049-016-0265-9
2. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012(6):CD000259. doi: https://dx.doi.org/10.1002/14651858.CD000259.pub3
3. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119(2):254-84. doi: 10.1037/0033-2909.119.2.254
4. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist* 2002;57(9):705-17. [published Online First: 2002/09/20]
5. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery* 2011;54(1):240-48.e1. doi: https://dx.doi.org/10.1016/j.jvs.2011.01.058

6. Veloski J, Boex JR, Grasberger MJ, et al. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher* 2006;28(2):117-28. doi: 10.1080/01421590600622665
7. Shute VJ. Focus on Formative Feedback. *Review of Educational Research* 2008;78(1):153-89. doi: 10.3102/0034654307313795
8. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education* 2016;50(9):933-42. doi: https://dx.doi.org/10.1111/medu.13063
9. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 1997;19(1):40-44. doi: 10.3109/01421599709019346
10. Sargeant J, Mann K, Sinclair D, et al. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41(6):583-91.
11. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17(1):15-26. doi: https://dx.doi.org/10.1007/s10459-011-9290-7
12. Xeroulis GJ, Park J, Moulton CA, 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(4):442-9. doi: https://dx.doi.org/10.1016/j.surg.2006.09.012
13. Li Q, Ma EL, Liu J, et al. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher* 2011;33(10):e549-55. doi: https://dx.doi.org/10.3109/0142159X.2011.600360
14. Elliott SN. Educational psychology: effective teaching, effective learning. 3rd ed. Boston: McGraw-Hill 2000.
15. Wadsworth BJ. Piaget's theory of cognitive and affective development: Foundations of constructivism. 5th ed. White Plains: Longman Publishing 1996.
16. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 1998;10(3):251-96. doi: 10.1023/A:1022193728205
17. 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 Medical Simulation* 2016;11(4):257-63. doi: https://dx.doi.org/10.1097/SIH.0000000000000165
18. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia* 2004;181(6):327-8.
19. Fitts PM, Posner MI. Human Performance. Belmont CA: Brooks/Cole 1967.
20. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine* 2008;15 doi: 10.1111/j.1553-2712.2008.00227.x
21. 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 ed. Oxford: Wiley Blackwell 2014:7-29.
22. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227-68. doi: 10.1207/s15327965pli1104_01

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	P1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	P2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	P5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	P7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	P7
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	P8
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	P8
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	P7-8
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	P9
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	P9, Tables 1 & 2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	P9-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	P10

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	P10-11

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	P9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	P10-11
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	P11 & Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	P12 & 19-20 Tables 1 & 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	P18, & 20 & Figures 2 & 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures 5 & 7
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	P18-19 Figure 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Figures 2 & 3
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	P30-31
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	P3, 34

PRISMA 2009 Checklist

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	P31
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	P36

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.