Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review

J M Monica van de Ridder,[1] William C McGaghie,[2] Karel M Stokking[3] & Olle T J ten Cate[4]

CONTEXT Feedback is considered important in medical education. The literature is not clear about the mechanisms that contribute to its effects, which are often small to moderate and at times contradictory. A variety of variables seem to influence the impact of feedback on learning. The aim of this study was to determine which variables influence the process and outcomes of feedback in settings relevant to medical education.

METHODS A myriad of studies on feedback have been conducted. To determine the most researched variables, we limited our review to meta-analyses and literature reviews published in the period from January 1986 to February 2012. According to our protocol, we first identified features of the feedback process that influence its effects and subsequently variables that influence these features. We used a chronological model of the feedback process to categorise all variables found.

RESULTS A systematic search of ERIC, PsycINFO and MEDLINE yielded 1101 publications, which we reduced to 203, rejecting papers on six exclusion criteria. Of these, 46 met the inclusion criteria. In our four-phase model, we identified 33 variables linked to *task performance* (e.g. task complexity, task nature) and *feedback reception* (e.g. self-esteem, goal-setting behaviour) by trainees, and to *observation* (e.g. focus, intensity) and *feedback provision* (e.g. form, content) by supervisors that influence the subsequent effects of the feedback process. Variables from all phases influence the feedback process and effects, but variables that influence the quality of the observation and rating of the performance dominate the literature. There is a paucity of studies addressing other, seemingly relevant variables.

CONCLUSIONS The larger picture of variables that influence the process and outcome of feedback, relevant for medical education, shows many open spaces. We suggest that targeted studies be carried out to expand our knowledge of these important aspects of feedback in medical education.

Medical Education 2015: 49: 658–673
doi: 10.1111/medu.12744

Discuss ideas arising from the article at www.mededuc.com discuss.

[1]Department of Education, Albert Schweitzer Hospital, Dordrecht, the Netherlands
[2]Department of Medical Education, Feinberg School of Medicine Northwestern University, Chicago, Illinois, USA
[3]Department of Educational Sciences, Utrecht University, Utrecht, the Netherlands
[4]Center for Research and Development of Education, University Medical Centre Utrecht, School of Medical Sciences, Utrecht University, Utrecht, the Netherlands

Correspondence: J M Monica van de Ridder, Centre for Research and Development, Albert Schweitzer Hospital, Albert Schweitzerplaats 25, Postbus 444, 3300 AK Dordrecht, the Netherlands.
Tel: 00 31 636 103294; E-mail: m.van.de.ridder@asz.nl

INTRODUCTION

Feedback is meant to improve employees' and learners' performance and to implement procedures. Feedback in clinical education may be defined as: 'Specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance.'[1] The literature on feedback is abundant.[2–6] Several studies show the impact of feedback on motivation and performance.[4–8]

In health care education, medical students receive feedback in the clinical context, a dynamic learning environment in which there is a lack of continuity and in which the relationships among patient, clinician and student are key.[9,10] Feedback is often provided by different clinicians who are usually not trained to do this, and allocating time to the provision of feedback is not common practice.[10,11] Feedback is a teaching and learning tool, with high educational impact.[10,12–15]

As feedback is considered a cornerstone in clinical education,[16] it is important that clinicians understand the feedback process and learn which variables relate to assessment and feedback procedures and to the effects of feedback. This knowledge can also support the skills of clerks and residents as feedback recipients. These understandings together should raise the quality of feedback procedures. Feedback can then be used optimally and will result in better collaboration, more competent physicians and better patient care.

The effects of feedback are equivocal and confusing[4] because feedback can both increase and decrease motivation and performance.[8] Whereas Hattie and Timperley report large overall effect sizes,[5] Kluger and DeNisi,[8] and Ivers *et al.*[17] report that the impact of feedback is small to moderate. One explanation for these inconsistent findings may refer to the existence of variables that influence the process and effect of feedback.

The feedback process can be considered a communication process of a looped nature, consisting of five phases[6,18–22] as depicted in Fig. 1.

In phase A_1, the feedback recipient (FR) receives instructions and performs a task according to certain standards. Phase B relates to the feedback provider's (FP) observation and interpretation of the FR's task performance, which the FP compares with an explicit or implicit standard. During phase C, feedback is communicated towards the FR. In phase D, the FR receives the feedback and interprets it.

When the same task is performed again, the cycle is closed (phase A_2). The feedback effect (ΔA_2–A_1) can be found when performances A_1 and A_2 are compared. Effects of feedback may be found in changes in the FR's cognition, attitude and performance that occur in consequence to the feedback received.

Figure 1 visualises the chronology of the feedback process in a schematic model. In reality, phases A and B will generally take place simultaneously, as will phases C and D. According to this model, phase B, the assessment phase, in which the observation and rating take place, directly influences phase C, in which the feedback is communicated. For example, if the same performance is observed by two clinicians independently, of whom one has observed the clerk many times and the other only once, although these

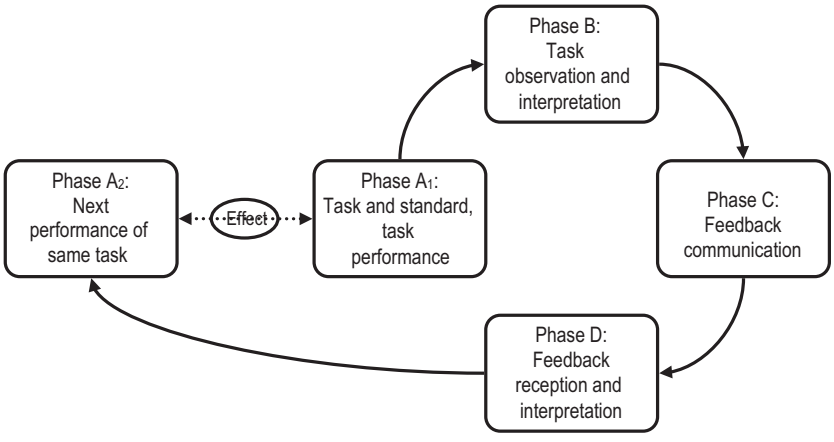

Figure 1 Representation of feedback process phases A_1, B, C and D, and the feedback effect, which becomes apparent when two performances are compared (ΔA_2–A_1).

clinicians witness the same performance, differences in their respective frames of reference may cause their observations to differ. This may result in two different feedback messages based on the same observation. When inter-rater reliability between two FPs observing a task appears low, and these observations are used for feedback, the FR will receive two different messages, which makes corrective action difficult. (For a more elaborate explanation, refer to Appendix S1, section I, online). Each phase can influence all subsequent phases. The phases A, B, C and D together influence the effect of feedback.

Feedback and models of feedback stem from a technical background[21] and over time feedback has come to be applied in various disciplines, such as communication, teaching and learning.[1] This feedback process model is based on Johnson's and Lasswell's models of communication.[23–27] It has similarities to other feedback models used in the communication, teaching and learning contexts.[2,3,28–31]

The aim of our study is to provide an overview of variables that influence the process of feedback, the effects of feedback, or both, and to describe how these influences relate to the different phases of the feedback process and the feedback effect. Cross-disciplinary items of literature from education, psychology, labour and management studies, which take both the process and the effects of feedback into account, were sourced. From this overview, directions for further research into feedback, both within and outwith health care education, can be derived.

We aimed to answer the following research questions: which variables in the feedback process influence either the first or subsequent phases in this process or a second performance on the same task? What is the direction of the influence of each variable on the subsequent phases of the feedback process or on the final effect?

METHODS

This article was prepared using most of the reporting conventions described in the PRISMA (*p*referred *r*eporting *i*tems for *s*ystematic reviews and *m*eta-*a*nalyses) Statement for meta-analyses and systematic reviews.[32]

Study eligibility and identification

To determine the most researched variables, we limited our source material to literature reviews and meta-analyses. The advantage of reviews and meta-analyses over individual studies is two-fold: they summarise common themes and conclusions, and they focus on a usually well-defined domain. This reduces the likelihood of reporting atypical or misleading results.[33]

We conducted searches of the ERIC (Educational Resources Information Center), PsycINFO and MEDLINE databases for meta-analyses and literature reviews published in English in peer-reviewed journals during the period from January 1986 to February 2012 (Fig. 2). In order to identify relevant articles as fully as possible, the search strategy was developed and discussed with two experienced reference librarians (LO'D and RK).[34] According to the database, we used the following thesaurus terms and medical subject headings: 'Reinforcement (psychology)'; 'Reinforcement, Verbal'; 'Reinforcement, Social'; 'Formative-Evaluation'; 'Personnel-Evaluation'; 'Student-Evaluation'; 'Job-Performance'; 'Informal-Assessment'; 'Performance-Based-Assessment'; 'Employee-Performance-Appraisal'; 'Error-Correction'; 'Feedback (psychological)'; 'Performance-Tests'; 'Knowledge-of-Results (psychology)', and 'Delayed-Feedback'. These were combined with 'State-of-the-Art-Reviews', 'Literature-Reviews' and 'Meta-Analysis'. To retrieve results pertaining to an educational and psychological context, our search of MEDLINE included terms related to medical education ('Schools, Medical', 'Students, Medical', 'Clinical Competence', 'Internship and Residency', 'Clinical Clerkship', 'Teaching', 'Education', 'education' [Subheading], 'Hospitals, Teaching', 'Competency-Based Education', 'Education, Continuing', 'Education, Medical, Continuing', 'Competency-Based Education', 'Education, Medical, Undergraduate' OR 'Education, Medical, Graduate').

One rater (JMMvdR) excluded articles according to one or more of the exclusion criteria (below) based on the information provided by abstracts and titles. Three raters (JMMvdR, WCM and KMS) independently rated the remaining articles according to the three inclusion criteria. Agreement on inclusion was reached by consensus.

Exclusion and inclusion criteria

We used six criteria to exclude articles based on their titles and abstracts. These were: (i) a non-relevant population (applied to studies based on animal data or on children, disabled persons or patients); (ii) non-behavioural topics (applied to studies focusing on descriptions, reviews, comparisons and

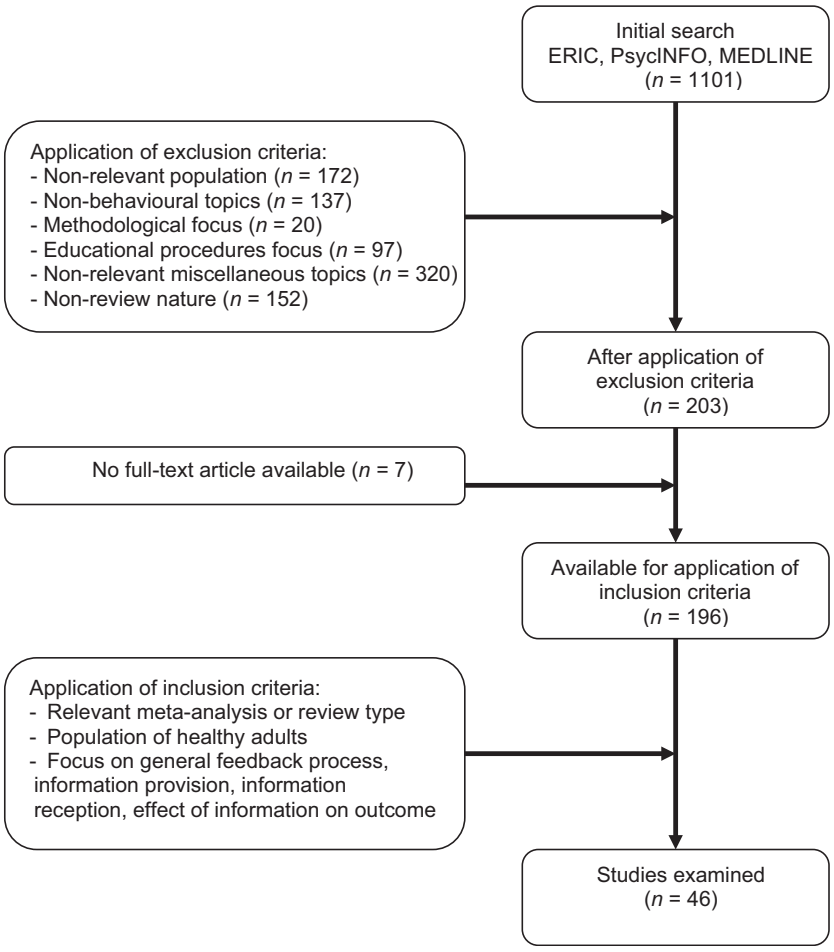

Figure 2 Search strategy and results.

evaluations of 'non-behavioural' topics, such as [student] feedback on teaching/learning programmes, materials, guidelines, curricula, etc.); (iii) methodological focus (applied to studies that focused on the methodology of meta-analyses and reviews in particular or on methodology problems in general); (iv) educational procedures focus (applied to studies focusing on differences in methods [e.g. traditional–authentic], models, measurement, tools and approaches of assessments and evaluation of performances and competencies); (v) non-relevant miscellaneous topics (applied to studies focused on topics other than feedback, such as job commitment and turnover, personnel selection, grading and scoring, self-assessment, influences or predictions on job or study performance, physical processes such as biofeedback, video feedback and rewards), and (vi) a non-review nature (applied to studies that provided comments or critique on reviews, or studies in which reviewing the literature was not the main purpose of the study).

Subsequently, from among all articles retained after the exclusion process, we included all studies that met all of the following criteria: (i) the use of meta-analysis or a systematic literature review of empirical research as the methodology, and the demonstration of good quality research by the reporting of search terms, time span and sources (the use of predetermined inclusion criteria and especially exclusion criteria was also considered a criterion for quality[35]); (ii) the use of a population of healthy adults, and (iii) a focus on the feedback process in general, or the process of providing information, or the process of receiving information, or the effect of the information on outcomes.

Process of selection of variables and synthesis of results

We used an eight-step protocol to select variables and synthesise findings. We applied the descriptive model (Fig. 1) as a tool to relate the variables to

the different phases in the feedback process (Appendix S1, Section II provides more details).

In step 1, we categorised each study as referring to one or more phases of the feedback process or the feedback effect. We first determined the dependent variables or the outcome variables of the included studies (Table S1, column H, online).

In step 2, we determined dependent, independent, moderator, mediator and confounding variables (Table S1, column I). Moderator, mediator and confounding variables elucidate the causal process by which the independent variable influences the dependent variable.

In step 3, we assessed the unit of measurement for each study. In meta-analyses, effect sizes are the usual units of measurement. Effect size *d* (ES*d*) indicates the extent to which means differ. Effect size *r* (ES*r*) indicates the size of correlation effects between variables. In literature reviews, units of measurement may be the number of studies that report a change, or the percentage of studies in which effects were found. Narrative reviews just describe differences. We reported only variables that were clearly indicated as effective.

In step 4, we identified and selected those variables which have a reported impact on the feedback process and feedback effect. In meta-analyses, we considered only variables with a significant effect. In quantitative and narrative reviews, we included only those variables that were clearly reported as influencing the direction of the effect.

In step 5, we established variables that did not influence the feedback process or the feedback effect or both (Table S1, column J).

In step 6, to detect which phases of the feedback process and feedback effect were influenced, we considered (i) which dependent variable was influenced, and (ii) in which phase of the feedback process or the feedback effect this variable could be classified.

In step 7, to create a comprehensive overview, we clustered variables with similar meanings or similar content (Table 1).

In step 8, from the variables that reportedly influenced the phases of the feedback process or the feedback effect or both, we determined the direction of the effect. Table S1 describes the characteristics of the included studies.

RESULTS

Studies identified

Figure 2 summarises the results of the search strategy. An initial search yield of 1101 publications was reduced to 46 relevant studies to be examined: 22 meta-analyses and 24 literature reviews. Dependent variables were related to phase A in one study, to phase B in 13 studies, to phase C in four studies, to phase D in two studies and to the feedback effect in 24 studies. The independent variables of six studies pertained to a combination of the phases and the effect (Table S1, column E). Ten of the 46 studies were performed in a health care setting (Table S1, column G).[14,16,36–43]

Section IV in Appendix S1 provides a detailed description of the evidence found in the literature reviews and meta-analyses about how each of the 33 variables influences the feedback process, the feedback effect, or both.

Variables identified

Based on the 46 reviews and meta-analyses, we found 33 variables related to task, standards and task performance (phase A_1), task observation and interpretation (phase B), feedback communication (phase C), and feedback reception and interpretation (phase D), influencing the feedback process, the feedback effect, or both. A description of each variable can be found in Table 1.

Phase A: general context, task, standard and first task performance by the FR

The nine variables that relate to phase A_1 are mostly about the task and include the number of different tasks that need to be performed,[44] complexity,[8,36,45–50] nature,[8,50–53] subject matter,[44] and the perception of the task.[45] One variable is about the organisation's culture, or the context in which the task will be performed.[47,49,54] Three variables relate to the learner who performs the task and who will receive feedback, and refer to, respectively, the FR's cultural background,[54–56] age[57] and initial skills.[36, 44, 45]

Phase B: observation, interpretation and rating by the FP

Twelve independent variables are embedded in the phase of observation, interpretation and rating, as

*Table 1 Descriptions of the independent variables that influence the feedback process, the feedback effect, or both**

Variable	Description
Task variation	Differences in level, subject matter, complexity of tasks performed during observation and rating
Task complexity	Extent to which a task is easy or difficult to analyse, to understand or to perform
Task nature	Basic or inherent features, or characteristics of a task (e.g. communication, memory, algorithm)
Task perception	Interpretation or impression based on one's understanding of the performed task
Task subject matter	Topic dealt with, or the subject represented in the task (e.g. mathematics, language, biology)
Culture and context	Circumstances that form the setting for the feedback process, in terms of which it can be fully understood (e.g. habits and customs of a country, or the working environment)
FR's age	Length of time the learner has lived
FR's cultural background	FR's previously acquired understanding or knowledge that allows utterances, beliefs and actions to have explicit meaning, especially in relation to cultural interpretation
FR's skills	Ability of the FR to perform the task
Training content and method	Topics contained in the training (e.g. rating errors, rating accuracy, rating instruments, etc.) Procedure for approaching the goals of the training: through discussion, lectures, practice, observation, etc.
Instrument	The tool or implement to observe and rate the FP's performance: mini-CEX, checklists, etc.
Assessment method (utility)	Usefulness of methods for measuring the impact of a course or lesson on learners' levels of attainment: observation, OSCEs, knowledge tests, etc.
Rubrics	Written instructions provided for candidates as part of a test
Purpose of observation	Reason why the observation is made (e.g. research or administration)
Focus of observation	Centre of interest during the observation
Intensity of observation	Degree or strength of the observation: in real life, based on video tapes, focused on specific performances, one to one, in a group
Type of standard	Category of the required or agreed level of quality or attainment used for the FR's observation or rating: result-oriented measures of performances, amount of time spent on a task, subjective measures of performances with space for personal interpretation
Time to build relationship	Effort put in to connecting with somebody else
FP's position	Particular way in which the FP is placed, especially as it influences the FP's power to act (e.g. subordinate, supervisor)
FP's task familiarity	Extent to which the FP has encountered or experienced a task
FP's cultural background	FP's previously acquired understanding or knowledge that allows utterances, beliefs and actions to have explicit meaning, especially in relation to cultural interpretation
Source of feedback	Person or thing from which the feedback message originates or can be obtained (e.g. computer, peers, colleagues)
Medium of feedback	Means by which the feedback message is communicated or expressed
Form of feedback	Style, design and arrangement of the feedback message as distinct from its content (examples of form: dialogue, as part of education, multifaceted interventions, organised feedback meeting)
Content of feedback	The material, information dealt with in the feedback message, as distinct from its form or style
Complexity of feedback	The extent to which the feedback is easy or difficult to analyse or to understand (e.g. difficult formulation, lengthy messages)
Timing of feedback	A particular point or period of time when feedback is given (e.g. immediate, delayed, after an item, after a test)
Frequency of feedback	The rate at which the feedback occurs over a particular period of time

Table 1 (Continued)

Variable	Description
Intensity of feedback	The degree or strength of the feedback communication (examples of aspects of intensity: over a long time period, by credible feedback providers, with rich content, with regular checks whether or not the message is understood, length of the feedback messages)
Time: feedback performance$_2$	Duration of interval between receipt of feedback and the next performance
Activities: feedback–performance$_2$	Tasks performed in the interval between receipt of feedback and next performance
FRs' self-esteem	FR's confidence in own worth or abilities
FRs' goal-setting behaviour	FR's ability to strive for an aim or desired result

* Descriptions of variables are based on definitions in the Oxford English Dictionary[89].
FP = feedback provider; FR = feedback recipient; mini-CEX = mini clinical examination; OSCE = objective structured clinical examination.

carried out by the FP before the communication of the feedback. A first group of variables is about the use of observation and rating methods and instruments, and includes training content and method,[58–60] instrument,[37,46,60–65] utility of assessment method,[14,38,39,53] and rubrics.[66] A second group pertains entirely to the mode of observation: the purpose (why),[46,61,62] the focus (what),[55,56,63] the intensity (how),[37,45,46,60,62] and the standard which is used during the observation.[61,63,64] The variable about the time available to the FP to build a relationship with the FR[61,67,68] links with the variable of the FP's position.

Three variables describe characteristics of the FP: the FP's position with respect to the FR;[50,51,60,62,65,68] the FP's acquaintance with the observed task,[50,51,65,68]and the FP's cultural background.[54]

Phase C: communication of the feedback message by the FP to the FR

The eight variables found in this category are about the person providing the feedback (source),[16,47] about the means (medium),[8,69] about the message form,[16,36,40,41,45,52,57] the message content,[4,8,49,52,70–73] the message complexity,[4] organisational aspects (timing,[4,52,72,74,75] frequency[8,42]) and intensity.[16,36,47,57]

Phase D: reception, perception and interpretation of the feedback message by the FR

Two variables pertain to the episode after the receipt of feedback and before the next performance (time interval: feedback to second performance[57,74] and disturbing activities in the period between the feedback and the second performance[74]). The other two variables relate to the learner who receives the feedback and refer to the FR's self-esteem and the FR's goal-setting behaviour.[8]

Direction of variables

To answer the second research question, we will first consider the phases of the process and the feedback effect that are influenced by the 33 variables, and then explain the influences of these variables (Table 2).

The main effect of the provision of feedback is that feedback is effective and improves performance,[8,16,36,40,70,76,77] such as in safety-related performance,[76] work productivity,[49] judgement abilities,[70] learner's goal-setting abilities,[4] and clinicians' and physicians' performance.[16,40] The impact of feedback is often small to moderate.[8,36,43,47,52]

Three variables influence the general context: task; standards, and first task performance (Table 2, phase A). Sixteen variables influence the observation, rating and interpretation by the FP (Table 2, phase B). The columns for phases C and D in Table 2 are almost empty. We found only two variables (training content and method, and the use of rubrics) that influenced the communication of the feedback message (Table 2, phase C). The variables 'culture and context', 'feedback medium' and 'feedback content' influence how the feedback is received, perceived and interpreted (Table 2, phase D). Twenty variables were found to influence the feedback effect (Table 2, Effect). It is striking that only two variables are related to observation, interpretation and rating.

Table 2 Literature sources that report the existence of effects of various variables within phases A–D on one or more subsequent phases in the feedback process*

Independent variables	Phase A	Phase B	Phase C	Phase D	Effect
Phase A: tasks, standards and first task performance					
Task variation	9				
Task complexity		4,17			6,12,15,19,25,31†
Task nature		4,22			2,11,12,16
Task subject matter	9				
Task perception					6
Culture and context		26		26	15,19,26
FR's cultural background		5,13,26			
FR's age					16
FR's skills	9				6†,31
Phase B: observation and interpretation					
Training content and method		40,45,46	40,45,46		
Nature of the instrument		3,7†,8,18,27,45			
Assessment method: utility		28,36			11,35
Rubrics		33	33		
Purpose of observation		17,18,32			
Focus of observation		3,5,13			
Intensity of observation		17,32,45			6,27
Type of standard		3,8,18			
Time to build relationship, trust, safety		18,38,43			
FP's position		4,7,22,32,38,45			
FP's task familiarity		4,7,22,38			
FP's cultural background		26			
Phase C: communication of the feedback message					
Feedback source					15,44
Feedback medium				29	12
Feedback form					2,6,16,31,37,42,44
Feedback content				20,39	2,10,12,19,23,30,39
Feedback complexity					39†
Feedback timing					2,14,21,30,39
Feedback frequency					12,41
Feedback intensity					15,16,31,44
Phase D: reception and interpretation of the feedback message					
Time: feedback–performance$_2$					16,21†
Activities: feedback–performance$_2$					21
FR's self-esteem					12
FR's goal-setting behaviour					12

* See Table S1 for further details of these reviews or meta-analyses.
† Results on these variables seem to be influential, but also show contra-indications.
FP = feedback provider; FR = feedback recipient.

Examples from each phase show variables described in meta-analyses or literature reviews as influencing the feedback process (Table 2).

The effects of 12 variables were unequivocal and showed a clear direction. Six variables influence outcome variables related to phase B observation, inter-

Table 3 Overview of variables that have a clear direction and an unequivocal effect on the observation, interpretation and rating, and the feedback effect

Influencing variable	Effect	Outcome measure
Observation, interpretation and rating		
FP rates *high complexity tasks*	→ Decrease in	Inter-rater agreement
FP has *high task familiarity*	→ Increase in	Agreement in ratings
FP is *trained* in using observation instruments	→ Decrease in	Rating errors
FP uses *rubrics*	→ Increase in	Reliability of scoring
FP and FR have *similar cultural background*	→ Higher	Performance ratings
FP has *time to build relationship* with FR	→ Higher	Correlations between subjective and objective performance measures
Feedback effect		
FR has *low initial task performance*	→ High	Feedback effect
Feedback message *threatens FR's self-esteem*	→ Low	Feedback effect
FR shows *goal-setting behaviour*	→ Increase in	Feedback effect
Feedback is part of a *multifaceted intervention*	→ Increase in	Feedback effect
Feedback content: *encouraging, specific, elaborate*	→ Increase in	Feedback effect
Feedback message is given *frequently*	→ Increase in	Feedback effect

FP = feedback provider; FR = feedback recipient.

pretation and rating, and six variables influence the feedback effect. Table 3 summarises our main findings.

Variables influencing the outcome of phase B: observation, interpretation and rating

Rating tasks of high complexity results in low inter-rater agreement

Task complexity influences how observers rate observed performances in the workplace.[46,50] Feedback providers show less agreement on ratings of highly complex tasks compared with those of tasks of low complexity. Possible explanations for this refer to the number of subtasks included in complex tasks, a lack of clear standards (for each subtask), that an FP is not skilled enough to perform the task him- or herself, and the difficulty of observing complex task behaviour.

High task familiarity in the FP leads to more agreement in ratings

Feedback providers who are familiar with the observed task show high inter-rater agreement. Inter-rater agreement in ratings of management tasks, diagnosis, job knowledge and judgement ability is higher among supervisors than among peers or subordinates. Supervisors have an overview of the entire task as a result of their competence or their experience in performing similar tasks themselves. However, peers show higher inter-rater agreement in ratings of observed interpersonal behaviour. The fact that peers work more often with their colleagues gives them better insight than subordinates or supervisors into these particular aspects.[50,51,65]

Training FPs in using observation instruments reduces rating errors

Rating errors influence the quality of the observation and, in a later phase, the content of the feedback message. Examples of such errors include the halo effect, the leniency effect, rating accuracy and observational accuracy.[58–60] Training in how to work with observation instruments reduces rating errors and increases accuracy.[60] Training in behaviour observation had a high positive impact on the quality of the task observation ($d = 0.59$).[58] Training in which discussion about the meanings of rating cate-

gories, practice and feedback are central components is very effective: it reduces rating errors by 50–85%.[59]

Use of rubrics increases the reliability of scoring and facilitates feedback communication

A rubric is: '. . .a scoring tool for qualitative ratings of authentic or complex student work. It includes criteria for rating important dimensions of performance as well as standards of attainment for those criteria.'[66] In rubrics, criteria and standards are made explicit so that the meaning and focus of dimensions of the task are clear, which results in high reliability of scoring.

Because of their concrete nature, FPs perceive rubrics as also facilitating feedback communication. Jonsson and Svingby did not use empirical data to illustrate their findings.[66]

Similarities in the cultural backgrounds of the FR and FP result in higher performance ratings

Learners' and supervisors' perceptions about acting as a professional are determined by (country) cultural background and by the organisational culture. An FR working in his own country and being supervised by an FP from the same cultural background achieves higher task performance ratings.[54] Because of the similarities in their cultural backgrounds, these individuals have comparable perceptions about professionalism and value certain job dimensions similarly.[54]

Making more time available for the FP to build a relationship with the FR leads to higher correlations between ratings and objective performance measures

Objective outcome measures, such as time spent on a task and a person's production pace, are less prone to observation bias than ratings of human feedback providers. Norton found that in a peer feedback situation, the FP's time investment in the relationship with the FR influenced the correlation between the FP's performance rating and objective measures of performance ($r = 0.69$).[61] When the FP's time investment in the relationship was high, the correlations between the FP's ratings and objective ratings of performance were high.[61] The results suggest that getting to know the FR reduces rating errors of bias from the FP. Especially for students with lower self-esteem or high levels of test anxiety, knowing the supervisor represents a factor of trust.[68]

Variables influencing the feedback effect

Low initial skill in the FR results in high feedback effects

Research in the clinical setting shows that feedback on clinical practice was more effective with FRs who did not meet the required baseline of recommended practice than with FRs who did meet this requirement.[36]An FR with low skills has more opportunities for improvement and effects may become visible more quickly.

The FR's self-esteem influences the feedback effect

Self-esteem was not mentioned often in the reviews and meta-analyses. This may be because most of the variables studied relate to impersonal rather than personal aspects. When feedback is perceived as threatening a person's self-esteem, the effect of the feedback is lower.[8]

Having FRs show goal-setting behaviour increases the feedback effect

When an FR has a goal in mind and knows what he or she wants to accomplish, the received feedback might be used to reach that goal. Feedback is more effective when an FR sets a goal than when he or she does not.[8]

Feedback as part of a multifaceted intervention increases the feedback effect

The instructional content in which feedback is embedded, especially in a test situation, increases the feedback effect (Table 1 [form of feedback]).[52] The combination of a feedback intervention with other educational interventions (so-called multifaceted interventions), such as audits, supervision, educational outreach visits or education, increases the feedback effect.[16,36,40,41,52]

Encouraging, specific, elaborate feedback content increases the feedback effect

Tailoring the content of the feedback message to the desired feedback effect increases its effectiveness.[49] Feedback consisting of specific, relevant and encouraging information, such as correct answers (the standard), task information, additional explanation and information about performance in relation to previous trials, is more effective than shallow feedback, such as general information, 'right'/'wrong' information, or compliments.[4,8,52,70,72,73]

Frequent feedback increases the feedback effect

With simply trainable, short, visual and psychomotor tasks, frequently provided feedback increases the feedback effect more than infrequent feedback.[8] Soumerai *et al.* found a positive change in drug-prescribing practice when ongoing feedback reports about medical doctor's drug-prescribing behaviour were provided.[42] Being used to receiving frequent feedback does not appear to decrease the feedback effect.[49]

Twenty-one of the 33 variables show equivocal results: they influence the process and effect of feedback in one study, but not in other studies, or influences of other variables made interpretations about direction impossible, or extra information is needed to determine the direction (Appendix S1, section III). For example, research on inter-rater reliability between FPs of different positions reveals it to be higher among supervisors than between supervisors and subordinates.[50,51] Based on these results, we assume that the position of raters influences ratings, and that ratings of supervisors correspond more closely with ratings from other supervisors than with ratings from subordinates. However, this does not give us information on who is a more reliable rater.

Several studies show that the length of the feedback (Table 1 [intensity of feedback]) influences the feedback effect: messages of different lengths have different effects.[16,47] However, the results do not make clear whether long or short messages are better. Guidelines for the optimum length of feedback cannot be derived from this and are probably task-dependent.

DISCUSSION

We met the objectives of our study: cross-disciplinary feedback research is brought together, and this results in an overview of 33 variables influencing, moderating or interacting with either the feedback process or the feedback effect, or both.

Following our summary (Table 3), we conclude that the reliability of feedback information and feedback scores is high when it pertains to low-complexity or familiar tasks and when observers are trained and use rubrics. Feedback is more favourable when the FP and FR share a cultural background, and feedback messages correlate with objective performance findings when the provider and recipient know each other well.

Effects of feedback are strong when the FR has poor initial task performance, when the message does not contain any threat to the recipient's self-esteem, when the recipient shows goal-setting behaviour, and when the message is encouraging, specific, elaborate and frequently given.

With reference to the framework, our study shows that: (i) variables from each phase of the feedback process influence the next phase in the feedback process, and (ii) all four phases in the feedback process influence the feedback effect. The consequence for clinical practice is that we should not only focus on how well the FP *communicates* the feedback (phase C), but also on the performed task (phase A), the observation and rating (phase B), and the FR's reception and perception of the feedback (phase D).

We can illustrate this with an example. When a senior clerk receives feedback, its effect may be low if the clerk has had difficulty in taking a history from an elderly couple who disagree (high task complexity). The resident supervising the clerk has only just started training (lacks expertise) (A→B). Consequently, the resident is not able to describe the observed behaviour (unspecific message) (B→C), which leads the clerk to conclude he or she performed poorly (influencing self-esteem) (C→D). This example shows how these variables affect subsequent phases in the feedback process. The influences of these processes on the effect of feedback on learning and behaviour cannot be predicted. They may *increase* the effect because the clerk understands that specific information is needed to improve the history taking (A). They may *decrease* the feedback because the lack of concentration and expertise in the supervising resident (B) results in a non-specific message (C), or because the message is interpreted as a threat to the clerk's self-esteem (D).

We learn from this that the feedback process is crucial to the feedback effect. Focusing on only one phase has large consequences for the ecological validity of feedback studies. Thus, if a model is proposed to predict the feedback effect, the influences of the four phases should be represented: *variables influencing phase A + variables influencing phase B + variables influencing phase C + variables influencing phase D determine the feedback effect.* We assume that the influences in the difference phases will interact.

We also assume that the impact of some phases will be stronger than that of others. However, we need more evidence to predict how feedback effects work in combination.

Conclusions about variables with a clear direction are that the FP can increase the quality of an assessment by being familiar with the task, by training in the use of observation instruments, by using rubrics, and by building a relationship with the FR. The variables of phase B are important because their outcomes inform the feedback in phase C.

In phase C, FPs can also increase the effectiveness of feedback if they give it frequently, if their feedback has encouraging, specific and elaborate content, and if their feedback does not threaten the FR's self-esteem. Feedback is also more effective when it is part of a multifaceted intervention. Feedback effectiveness is improved when the FR is able to set goals.

Limitations

The present study has limitations that should be taken into account in considering the study and its contributions. This meta-review does not pretend to give an exhaustive overview of all the literature about feedback. Several studies and variables are missing for a variety of reasons. For example, 'new' or 'rare' topics, like feedback propensities,[78] feedback framing[79] and the characteristics of FPs[80] require more publications in order to be included in a qualitative review or meta-analysis. Sometimes these topics will turn up in narrative reviews.[60] Some reviews or meta-analyses were not identified by the particular search criteria we used, such as those by Sadler[81] and Black and Wiliam,[82] or because of particular methodological criteria, such as studies on the Barnum effect,[83] feedback seeking,[84] and the review by Hattie and Timperley,[5] which did not mention search criteria. Some studies were published outwith the time span we selected (January 1986 to February 2012), such as a review on the effect of providing feedback on the FP published in 1984,[85] and, finally, some variables may have been researched in single qualitative studies or experiments, but may not have been the subject of meta-analyses or reviews.

A consequence of focusing on meta-analyses and literature reviews is that the reported variables are not new; they have been researched in primary studies. Meta-analyses and reviews often suffer from threats to validity caused by the mixing of dissimilar studies, publication bias and the inclusion of poor quality studies.[86–88] This leads to misrepresentations of outcomes; we cannot exclude the possibility that this may apply to this meta-review.

We took the quality of the included studies into account, and reported only variables from narrative reviews that support the findings from other studies included, but this may still have influenced our results. Further, not all included variables should be given the same weight. Variables in meta-analyses and in systematic quantitative reviews deserve a higher weighting than those in narrative reviews because they are based on the inclusion of more studies.

The downside of bringing cross-disciplinary research together is that it involves a variety of study contexts, tasks and participants. Results found in one context cannot simply be generalised to other contexts: both task aspects and context influence findings.[8,47,52,54]

Research agenda

As Table 2 shows, we were able to identify very few relationships between variables related to observation, interpretation and rating (phase B) and the feedback effect (Effect column). Students often address the importance of observation before receiving feedback. Issues for further research might concern the nature of the relationship between being observed and the feedback effect, and how the reliability and validity of observation instruments influence the feedback process and the feedback effect.

Most variables influence phase A, phase B or the feedback effect. The large number of empty cells in the columns for phases C and D indicates a lack of systematic research into variables influencing the feedback communication (phase C) and the reception, perception and interpretation of feedback (phase D). Of the 46 studies included, three[38,60,68] and six[4,38,54,60,69,71] studies presented outcome variables that related to phases C and D, respectively (Table S1, column E). None of these studies explicitly pertained to feedback communication (phase C). Of the six studies relevant to phase D, only one focused explicitly on the reception, perception and interpretation of the feedback. This illustrates the complexity of researching these topics. For further research into the variables of phases C and D, we have three recommendations. Firstly, research might determine whether the variables listed in Table 2 influence these phases, in which direction and how much. Secondly, investigators might study further

aspects of verbal behaviour, such as the framing of feedback messages, the 'quality' of explanations and examples in the feedback, and the order of the positive and negative aspects of the feedback message. Lastly, research might explore aspects of non-verbal behaviour, such as tone of voice and facial expression, and search for features in the FP's and FR's non-verbal behaviour that might contribute to the better communication and reception of the feedback.

Five variables show mixed results in some phases (Table 2): they influence the feedback process and the effect, but they also show contra-indications. This indicates that more systematic research on this topic is needed.

Although the FP and FR are crucial actors in the feedback process, the numbers of independent variables do not reflect this. Good quality systematic reviews on the personal characteristics of FRs and FPs are lacking.[8] Personal characteristics are often not incorporated because of their complexity.[20,85] Further research on this topic is also suggested.

Practical implications

To make feedback effective, systematic research on the relationships among variables represented by empty cells will contribute to our knowledge of effective feedback; evidence-based guidelines on influential variables in different phases of the feedback process are needed to bring this knowledge into practice. Clearly, focusing only on the phase of *feedback communication* does not guarantee that feedback will be effective. A focus on all phases of the feedback process is likely to yield a more predictable effect.

Techniques that appear influential in other disciplines include the use of rubrics,[66] and systematic approaches such as the productivity measurement and enhancement system (ProMES), in which regular outcome-based feedback is provided on different indicators.[49]

Training in assessment and feedback can focus on variables that clearly affect the quality of observation, interpretation, rating, and the effectiveness of feedback. In addition to rater training, education about the development of multifaceted assessments in clinical environments in which feedback is embedded should receive attention. Courses on receiving feedback and goal setting may also increase the utility of feedback in clinical settings for clerks and residents.

Contributors: JMMvdR and WCM established the systematic search. JMMvdR, KMS and WCM selected the articles. All authors contributed to the analysis and interpretation of data. JMMvdR drafted the article. All authors contributed to the critical revision of the paper and approved the final manuscript for publication.
Acknowledgements: the authors thank Linda O'Dwyer and Ramune Kubilius, Galther Health Sciences Library, Northwestern University Chicago, IL, USA, for their help and valuable information on literature search strategies.
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 van de Ridder JMM, Stokking KM, McGaghie WC, ten Cate OTJ. What is feedback in clinical education? *Med Educ* 2008;**42**:189–97.
2 Ilgen DR, Barnes Farrell JL, McKellin DB. Performance appraisal process research in the 1980s: what has it contributed to appraisals in use? *Organ Behav Hum Decis Process* 1993;**54** (3):321–68.
3 Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings *Pers Psychol* 2005;**58** (1):33–66.
4 Shute VJ. Focus on formative feedback. *Rev Educ Res* 2008;**78** (1):153–89.
5 Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;**77** (1):81–112.
6 Cusella LP. Feedback, motivation and performance. In: Porter F, Jablin F, Putnam L, Roberts K, eds. *Handbook of Organizational Communication: An Interdisciplinary Perspective*. Thousand Oaks, CA: Sage Publications 1987;624–678.
7 Vollmeyer R, Rheinberg F. Motivational effects on self-regulated learning with different tasks. *Educ Psychol Rev* 2006;**18**:239–53.
8 Kluger AN, DeNisi A. Effects of feedback intervention on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;**119** (2):254–84.
9 Dornan T. Experience-based learning: learning clinical medicine in workplaces. PhD thesis; University of Maastricht, Maastricht 2006.
10 Irby DM. Teaching and learning in ambulatory care settings: a thematic review of the literature. *Acad Med* 1995;**70** (10):898–931.
11 Bowen JL, Irby DM. Assessing quality and costs of education in the ambulatory setting: a review of the literature. *Acad Med* 2002;**77** (7):621–80.
12 Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach* 2007;**29** (9):855–71.
13 McGee SR, Irby DM. Teaching in the outpatient clinic. Practical tips. *J Gen Intern Med* 1997;**12** (Suppl 2):34–40.

14 Hamilton KE, Coates V, Kelly B, Boore JRP, Cundell JH, McFetridge B, McGonigle M, Sinclair M. Performance assessment in health care providers: a critical review of evidence and current practice. *J Nurs Manag* 2007;**15** (8):773–91.
15 Irby DM, Wilkerson L. Teaching rounds. Teaching when time is limited. *BMJ* 2008;**336** (7640):384–7.
16 Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach* 2006;**28** (2):117–28.
17 Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O' Brien MA, Johansen M, Grimshaw J, Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;**6**:CD000259.
18 Clement DA, Frandsen KD. On conceptual and empirical treatments of feedback in human communication. *Commun Monogr* 1976;**43** (1): 11–28.
19 Frandsen KD, Millis MA. On conceptual, theoretical and empirical treatments of feedback in human communication: fifteen years later. *Commun Rep* 1993;**6** (2):79–91.
20 Ilgen DR, Fisher CD, Taylor SM. Consequences of individual feedback on behaviour in organisations. *J Appl Psychol* 1979;**64** (4):349–71.
21 Richardson GP. *Feedback Thought in Social Science and Systems Theory*. Waltham, MA: Pegasus Communications 1999.
22 Watzlawick P, Helmick Beavin J, Jackson DD. *Pragmatics of Human Communication. A Study of Interactional Patterns, Pathologies and Paradoxes*. New York, NY: WW Norton & Co. 1967.
23 Johnson W. *People in Quandaries. The Semantics of Personal Adjustment*. New York, NY: Harper & Row 1946.
24 Johnson W. Speech and personality. In: Bryson L, ed. *The Communication of Ideas. A Series of Addresses*. New York, NY: Cooper Square Publishers 1964;84–102
25 Lasswell HD. The structure and function of communication in society. In: Bryson L, ed. *The Communication of Ideas. A Series of Addresses*. New York, NY: Cooper Square Publishers 1964;37–51
26 Lasswell HD. Describing the contents of communications. In: Smith BL, Lasswell HD, Casey RD, eds. *Propaganda, Communication, and Public Opinion. A Comprehensive Reference Guide*. Princeton, NJ: Princeton University Press 1946;74–79
27 Lasswell HD. Describing the effects of communications. In: Smith BL, Lasswell HD, Casey RD, eds. *Propaganda, Communication, and Public Opinion. A Comprehensive Reference Guide*. Princeton, NJ: Princeton University Press 1946;95–117
28 Mory EH. Feedback research revisited. In: Jonassen DH, ed. *Handbook of Research on Educational Communications and Technology*. Mahwah, NJ: Lawrence Erlbaum Associates 2004;745–784
29 Gallagher JJ. Teaching and learning: new models. *Ann Rev Psychol* 1994;**45**:171–95.
30 Fenner DE. Avoiding pitfalls: lessons in surgical teaching. *Obstet Gynecol Clin N Am* 2006;**33** (2):333–42.
31 Spark JI, Rowe S. Clinical governance: its effect on surgery and the surgeon. *ANZ J Surg* 2004;**74** (3):167–70.
32 Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Ann Intern Med* 2009;**151** (4):264–9.
33 Harden A, Thomas J. Methodological issues in combining diverse study types in systematic reviews. *Int J Soc Res Meth* 2005;**8** (3):257–71.
34 Albanese M, Norcini J. Systematic reviews: what are they and why should we care? *Adv Health Sci Educ Theory Pract* 2002;**7** (2):147–51.
35 Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross-sectional study. *BMJ* 2005;**330** (7499):1053.
36 Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;**19**:CD000259.
37 Craig S. Direct observation of clinical practice in emergency medicine education. *Acad Emerg Med* 2011;**18** (1):60–7.
38 Rout A, Roberts P. Peer review in nursing and midwifery: a literature review. *J Clin Nurs* 2008;**17** (4):427–42.
39 Miller A, Archer J. Impact of workplace-based assessment on doctor's education and performance: a systematic review. *BMJ* 2010;**341**:c5064.
40 Steinman MA, Ranji SR, Shojania KG, Gonzales R. Improving antibiotic selection: a systematic review and quantitative analysis of quality improvement strategies. *Med Care* 2006;**44** (7):617–28.
41 Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet* 2005;**366** (9490):1026–35.
42 Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989;**67** (2):268–317.
43 Cheraghi-Sohi S, Bower P. Can the feedback of patient assessments, brief training, or their combination, improve the interpersonal skills of primary care physicians? A systematic review. *BMC Health Serv Res* 2008;**8**:179.
44 Huang C. Magnitude of task-sampling variability in performance assessment: a meta-analysis. *Educ Psychol Meas* 2009;**69** (6):887–912.
45 Fukkink NG, Trienekens N, Kramer LJC. Video feedback in education and training: putting learning in the picture. *Educ Psychol Rev* 2011;**23** (1):45–63.
46 Murphy KR, Herr BM, Lockhart MC, Maguire E. Evaluating the performance of paper people. *J Appl Psychol* 1986;**71** (4):654–61.

47 Li S. The effectiveness of corrective feedback in SLA: a meta-analysis. *Lang Learn* 2010;**60** (2):309–65.
48 Clariana RB, Koul R. Multiple-try feedback and higher-order learning outcomes. *Int J Instr Media* 2005;**32** (3):239–45.
49 Pritchard RD, Harrell MM, DiazGranados D, Guzman MJ. The productivity measurement and enhancement system: a meta-analysis. *J Appl Psychol* 2008;**93** (3):540–67.
50 Conway JM, Huffcutt AI. Psychometric properties of multisource performance ratings: a meta-analysis of subordinate, supervisor, peer, and self-ratings. *Hum Perform* 1997;**10** (4):331–60.
51 Viswesvaran C, Ones DS, Schmidt FL. Comparative analysis of the reliability of job performance ratings. *J Appl Psychol* 1996;**81** (5):557–74.
52 Bangert DR, Kulik CLC, Kulik JA, Morgan M. The instructional effect of feedback in test-like events. *Rev Educ Res* 1991;**61** (2):213–38.
53 Kingston N, Nash B. Formative assessment: a meta-analysis and a call for research. *Educ Meas* 2011;**30** (4):28–37.
54 Claus L, Briscoe D. Employee performance management across borders: a review of relevant academic literature. *Int J Manag Rev* 2009;**11** (2):175–96.
55 Kraiger K, Ford JK. The relation of job knowledge, job performance, and supervisory ratings as a function of ratee race. *Hum Perform* 1990;**3** (4):269–79.
56 Ford JK, Kraiger K, Schechtman SL. Study of race effects in objective indices and subjective evaluations of performance: a meta-analysis of performance criteria. *Psychol Bull* 1986;**99** (3):330–7.
57 Lyster R, Saito K. Oral feedback in classroom SLA. *Stud Sec Lang Acquis* 2010;**32** (2):265–302.
58 Woehr DJ, Huffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Organ Psychol* 1994;**67** (3):189–205.
59 Smith DE. Training programmes for performance appraisal: a review. *Acad Manage Rev* 1986;**11** (1):22–40.
60 Wanguri DM. A review, an integration, and a critique of cross-disciplinary research on performance appraisals, evaluations, and feedback, 1980–1990. *J Bus Commun* 1995;**32** (3):267–93.
61 Norton SM. Peer assessments of performance and ability: an exploratory meta-analysis of statistical artefacts and contextual moderators. *J Bus Psychol* 1992;**6** (3):387–99.
62 Jawahar IM, Williams CR. Where all the children are above average: the performance appraisal purpose effect. *Person Psychol* 1997;**50** (4):905–25.
63 Bommer WH, Johnson J, Rich GA, Podsakoff PM. On the interchangeability of objective and subjective measures of employee performance: a meta-analysis. *Person Psychol* 1995;**48** (3):587–605.
64 Heneman RL. The relationship between supervisory ratings and results-oriented measures of performance: a meta-analysis. *Person Psychol* 1986;**39** (4):811–26.
65 Harris MM, Schaubroeck J. A meta-analysis of self–supervisor, self–peer, and peer–supervisor ratings. *Person Psychol* 1988;**41** (1):43–62.
66 Jonsson A, Svingby G. The use of scoring rubrics: reliability, validity and educational consequences. *Educ Res Rev* 2007;**2** (2):130–44.
67 van Gennip NAE, Segers MSR, Tillema HH. Peer assessment for learning form a social perspective: the influence of interpersonal variables and structural features. *Educ Res Rev* 2009;**4** (1):41–54.
68 Sasanguie D, Elen J, Clarebout G, van den Noortgate W, Vandenabeele J, De Fraine B. Disentangling instructional roles: the case of teaching and summative assessment. *J High Educ* 2011;**36** (8):897–910.
69 Hepplestone S, Holden G, Irwin B, Parkin HJ, Thorpe L. Using technology to encourage student engagement with feedback: a literature review. *Res Learn Tech* 2011;**19** (2):117–27.
70 Balzer WK, Doherty ME, O'Connor R. Effects of cognitive feedback on performance. *Psychol Bull* 1989;**106** (3):410–33.
71 Tang SH, Hall VC. The overjustification effect: a meta-analysis. *Appl Cogn Psychol* 1995;**9** (5):365–404.
72 Jaehnig W, Miller ML. Feedback types in programmed instruction: a systematic review. *Psychol Rec* 2007;**57** (2):219–32.
73 Karelaia N, Hogarth RM. Determinants of linear judgement: a meta-analysis of lens model studies. *Psychol Bull* 2008;**134** (3):404–26.
74 Travlos AK, Pratt J. Temporal locus of knowledge of results: a meta-analytic review. *Percept Mot Skills* 1995;**80** (1):3–14.
75 Kulik JA, Kulik C. Timing of feedback and verbal learning. *Rev Educ Res* 1988;**58** (1):79–97.
76 McAfee RB, Winn AR. The use of incentives/feedback to enhance work place safety: a critique of the literature. *J Safety Res* 1989;**20** (1):7–19.
77 Azevedo R, Bernard RM. A meta-analysis of the effects of feedback in computer-based instruction. *J Educ Comput Res* 1995;**13** (2):111–27.
78 Herold DM, Fedor DB. Individual differences in feedback propensities and training performance. *Hum Resour Manage R* 2003;**13** (4):675–89.
79 Waung M, Jones DR. The effect of feedback packaging on ratee reactions. *J Appl Soc Psychol* 2005;**35** (8):1630–52.
80 Menarchy EP, Knight AM, Kolodner K, Wright SM. Physician characteristics associated with proficiency in feedback skills. *J Gen Intern Med* 2006;**21** (5):440–6.
81 Sadler DR. Formative assessment and the design of instructional systems. *Instr Sci* 1989;**18** (2):119–44.
82 Black P, Wiliam D. Assessment and classroom learning. *Assess Educ Princ Pol Pract* 1998;**5** (1):7–75.
83 Furnham A, Schofield S. Accepting personality test feedback: a review of the Barnum effect. *Curr Psychol Res Rev* 1987;**6** (2):162–78.

84 Madzar S. Feedback-seeking behaviour: a review of the literature and implications for HRD practitioners. *Hum Resource Dev Q* 1995;**6** (4):337–49.
85 Larson JR. The performance feedback process: a preliminary model. *Organ Behav Hum Dec* 1984;**33** (1):42–76.
86 Delgado-Rodriguez M. Systematic reviews of meta-analyses: applications and limitations. *J Epidemiol Commun Health* 2006;**60** (2):90–2.
87 Sharpe D. Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clin Psychol Rev* 1997;**17** (8):881–901.
88 Chambers EA. An introduction to meta-analysis with articles from the *Journal of Educational Research* (1992–2002). *J Educ Res* 2004;**98** (1):35–44.
89 Oxford English Dictionary. Oxford: Oxford University Press 2013. http://www.oxfordreference.com.proxy.library.uu.nl/browse?type_0=englishdictionaries. [Accessed 31 May 2014.]

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supplementary Materials in which the methodology and the effect sizes used in the meta-review are explained. Further it consists of a detailed description of the results in each phase of the feedback process.

Table S1. Overview of study characteristics of included meta-analyses (n=22) adn reviews (n=24) in alphabethical order.

Received 31 May 2014; editorial comments to author 29 July 2014, accepted for publication 4 February 2015