Grades in formative workplace-based assessment – a study of what works for whom and why

Janet Lefroy,[1] Ashley Hawarden,[1] Simon P Gay,[1] Robert K McKinley[1] & Jennifer Cleland[2]

Abstract

Context

Grades are commonly used in formative workplace-based assessment (WPBA) in medical education and training but may draw attention away from feedback about the task. The dilemma is that the self-regulatory focus of a trainee must include self-awareness relative to agreed standards, which implies grading.

In this study we aimed to understand the meaning which medical students construct from WPBA feedback with and without grades, and what influences this.

Methods

Year 3 students were invited to take part in a randomised crossover study in which each student was their own control. Each student had one WPBA with and one without grades, and then chose whether or not to have grades with their third WPBA. These preferences were explored via semi-structured interviews. A realist approach to analysis was used to gain understanding of student preferences and the impact of feedback with and without grades.

Results and discussion

Of students who had feedback with and without grades, 65 (78%) then chose to have feedback with grades and 18 (22%) without grades their third WPBA. 24 students were interviewed.

Students described how grades locate their performance and calibrate their self-assessment. For some, low grades focused attention and effort. Satisfactory and high grades enhanced self-efficacy.

Grades are also concrete, powerful and blunt, can be harmful and need explanation to help students create helpful meaning from them. Low grades risk reducing self-efficacy in some and may encourage others to focus on proving their ability rather than on improvement.

A metaphor of the semi-permeable membrane is introduced to understand how students reduced potential negative effects and enhanced the positive effects of feedback with grades by selective filtering and pumping.

Conclusion

This study illuminates the complexity of the processing of feedback by its recipients, and informs the use of grading in provision of more effective, tailored feedback.

Background

Feedback is an important teaching and learning tool (1–3). Indeed, workplace-based assessment (WPBA) of the competence and attitudes of trainees with feedback from their assessor is considered one of the most powerful interventions in medical education (2,4,5). For example, Veloski *et al.*, found that feedback had a positive impact on physician performance in 74% of 41 studies reviewed (5). However, there are sometimes powerful negative effects and problems with feedback in higher education (6). Students commonly report feedback as difficult to understand (7); non-specific (providing only vague and sweeping generalisations) (8); or difficult to act on (6,9). Overly critical feedback does not encourage learning, can be perceived as an attack and have a destructive effect (1,10–12).

Grades (e.g. below/meets/above expectations) have long been a component of WPBA in medical training (for example in the mini-CEX (13) and the Leicester Assessment Package (14)). This seems logical in a profession where both competence in clinical practice and knowing your own limits of competence are essential, and in which enabling learners to self-monitor in relation to competency requirements is an important goal (15,16). Yet there is a paradox: giving grades as part of a student's feedback in school settings and elsewhere in undergraduate education can reduce the effectiveness of feedback and may reduce student performance (1).

In research into what makes feedback effective, ineffective or harmful, the relationship between the recipient and giver, the state of mind and maturity of the recipient, the feedback process and the feedback content (11,12,17–19) are key factors. Feedback should clarify the student's position and progress relative to required goals, and suggest how to attain the goals (1); feedback is most useful if it is from a trusted assessor who knows the student (5) who is in a learning frame of mind (20) and there is a diagnostic and supportive dialogue between assessor and student which enhances motivation (21) because the suggestions in the feedback align with the student's goals and therefore seem relevant (11).

Factors which affect the impact of grades in formative feedback have been researched (8,10,22). While grading can give an unambiguous answer to a student's questions "Where am I relative to where I am going?" and "How well am I progressing?" which are the basis of future learning, and may increase involvement and effort, it may not improve performance because it mixes feedback about the person with feedback about the task thus distracting attention from feedback about the task (1,10). Grading may affect the tutor-student relationship because the tutor is both helping the individual to learn and passing judgement, and is thus explicitly in a position of power (5,8,23–25). Grades may also lack value in WPBA because differences in assessor calibration make standardisation of assessment almost impossible (26,27). This would be of particular concern if students were more interested in their grades than in the qualitative information which may accompany them.

So why do we use grades at all in formative WPBA? Should we stop? To date, most research on grading in formative assessment has been in school children and non-medical higher education. While useful for hypothesis generation (10), the findings may not be directly applicable to today's outcomes-based and outcomes-focused medical education. While there is a move towards portfolio-based self-directed learning without grading in some medical schools (28), medical students learning skills value grades (29). Does grading encourage or discourage feedback-seeking and self-directed learning? Or does this depend on something within the student, as suggested by research into the differing receptiveness to formative feedback depending on performance in summative assessment (30,31), when students who paid least attention to their feedback were those who had narrowly passed the assessment. Self-regulation theories suggest that within each of our students are two basic self-regulation systems which co-exist but may conflict (22,32). These two systems - the promotion (seeking good) and prevention (avoiding harm) approaches - may both be active in response to feedback. How does grading influence this balance? There is a clear need for research comparing the impact of feedback with and without grades within medical education, as a contribution to understanding how assessment influences learning (33,34).

In this study we aimed to understand first the meaning which medical students construct from workplace-based assessment and feedback with and without grades and, second, how this is influenced by the students' internal and external environment. A further aim was to use this information to develop more effective, individually-tailored feedback processes. We did so using a crossover study in which each student was their own control. We compared student experience of, and preference for, grades or no grades. A realist logic of analysis was used to gain understanding of student preferences and the impact of feedback with and without grades (35,36).

Context

Keele University School of Medicine has a five-year programme with repeated formative WPBAs of students' consultation skills using the Generic Consultation Skills (GeCoS) assessment tool (37,38). Each WPBA includes a face-to-face discussion between student and tutor about the student's strengths, areas which need to be improved and strategies for achieving this. Students receive a written summary of the discussion including their grades for each assessed domain of their consultations. The grading scale (Must Improve - Borderline – Proficient – Very Good) is referenced to the standard required of a graduating doctor (15). The first formal WPBAs take place during a four-week general practice (family medicine) placement at the end of Year 3 after the summative knowledge assessment and Objective Structured Clinical Examination (OSCE) when each student has three WPBAs - done in weeks one, two and four of this placement, and receives feedback from their GP tutor.

Methods

Recruitment

During the academic year of 2011-2012, all year three students and their GP tutors were invited by email to take part in the study. Participation was voluntary and non-participation had no influence on any aspect of teaching or progress.

Participation for students involved:

- Opting to receive grades for all three WPBAs (our practice at that time) or consenting to randomisation into one of two study groups:

  ➢ Group 1: First WPBA with grades, second without, third student's choice.

  ➢ Group 2: First WPBA without grades, second with, third student's choice.

- Consent (or not) for the research team to access their recent summative OSCE results (stations passed out of 12) and the three written WPBA summaries from the general practice placement.

- Consent (or not) to be interviewed in the four weeks following the placement. .

Participation for GP tutors involved:

- Consent for the research team to access their WPBA written feedback.

Students who consented to randomisation therefore experienced WPBA with and without grades before being asked their preference (grades or no grades) for the third and final assessment. This preference was obtained via an online questionnaire in the third week of their placement.

In order to sample the widest spectrum of student approaches to grading, maximal variation sampling was used when selecting interviewees with respect to gender, attainment in the recent OSCE and preference for each of the three options: grades with their third WPBA; no grades with their third WPBA; or grades with all three WPBAs. Semi-structured interviews were conducted at the end of the placement exploring what students felt about: their placement; their progress with consultation skills; their views on WPBA; what they could recall of each of the three WPBAs; why they chose grades or no grades and the impact of having or not having grades on their feedback; whether they agreed with their feedback and would wish to have such a choice in future WPBAs. One researcher (JL) conducted the interviews. Interviews were audio-recorded with consent, and transcribed verbatim.

Analysis

We chose a realist approach to data analysis, in which the question "What works for whom under what circumstances and why?" is asked. The aim of realist evaluation is to understand relationships between context and outcomes by discovering some of the workings or 'mechanisms' of the 'black box' of complex interventions (such as feedback with and without grades from a WPBA). These are termed CMO (context-mechanism-outcome) configurations (36,39). We considered realist evaluation to be closely aligned with the study aims: subjects were involved in the process of working out why (mechanism) they had reacted in the way they did (outcome) in different feedback situations (context). Furthermore, realist approaches can raise suggestions for adjustment of the complex intervention under study once the relationships between context, mechanisms and outcomes are better understood (36).

The realist evaluation approach involves testing an initial programme theory about the working of the 'black box'. A programme theory is a statement of what we think it is about the intervention which generates change. Our initial programme theory was based on two assumptions from the literature. Firstly that feedback should answer the student's questions "Where am I going?", "How am I progressing?" and "How can I make progress?" which should result in discovering strategies for improvement (1), and that adding grades to the context would enhance this outcome by adding clarity. Secondly, based on theories of self-regulated learning, in which goal level, persistence, effort, and self-efficacy are the self-regulation constructs with the strongest effects on learning (21,40), that the motivational effects of grading would also have an important influence on the outcome. We were expecting these motivational effects to be either positive or negative depending on both the external context (feedback being positive or negative) and the student's multi-faceted internal context, and that mechanisms would include confirmation or conflict with self-assessment; desire to improve and desire to avoid failure (10).

The initial programme theory to be tested against our data and refined was therefore that if the student has a trusted assessor (external context) and a learning goal approach (internal context), they will find that grades (external context) clarify (mechanism) and energise (mechanism) their efforts to find strategies to improve (outcome). If students were more performance oriented (internal context) they would find that satisfactory grades (external context) reassured (mechanism) and therefore reduced efforts to improve (outcome). We were less sure what to expect to find as the outcome of lower than expected formative grades, and were looking for outcomes and their explanations.

The transcripts were therefore examined for context-mechanism-outcome (CMO) configurations – what effect (outcome) did the feedback with and without grades have (contexts), what caused these effects (mechanisms) and in what internal and external learning environment (contexts) did these occur.

The first two transcripts were analysed by all authors to develop our joint understanding of what constitutes a context, mechanism and outcome in formative WPBA. A table was produced for each transcript listing the CMO configurations identified, with columns for student comments about feedback with and without grades (the manipulated variable in the context). Subsequent transcripts

were coded separately by JL and AH who compared their analyses. Where interpretation was difficult, one or two of the other researchers also analysed the transcript to reach a consensus.

The authors then compared CMO configurations containing cognitive, self-regulatory and other explanations of the effects of feedback with and without grades, seeking evidence to corroborate and refine our initial programme theory. Where it was not corroborated, alternative explanations were sought where different mechanisms might have been operating in different contexts.

Reflexivity

The research team comprised stakeholders in the programme under study (JL SPG RKM AH) in collaboration with JC from another UK medical school who was invited to provide an external perspective. JC is a clinical psychologist who has worked in medical education for 14 years. Her interests in feedback stem from working as a medical educator in undergraduate medicine, and from researching "failure to fail"(25). JL SPG and RKM are clinicians who were involved in the development of the instruments for, and the programme of WPBA and feedback. They therefore conducted analysis from that stakeholder viewpoint but with a genuine lack of certainty about whether feedback should or should not include grades, AH was a Keele medical student who at the time of the study had recently completed Year 3 and therefore brought the student perspective to analysis. As JL was a clinical tutor known to the interviewees, we were aware that this may influence participant responses so efforts were made to diminish this effect in the preamble to interviews.
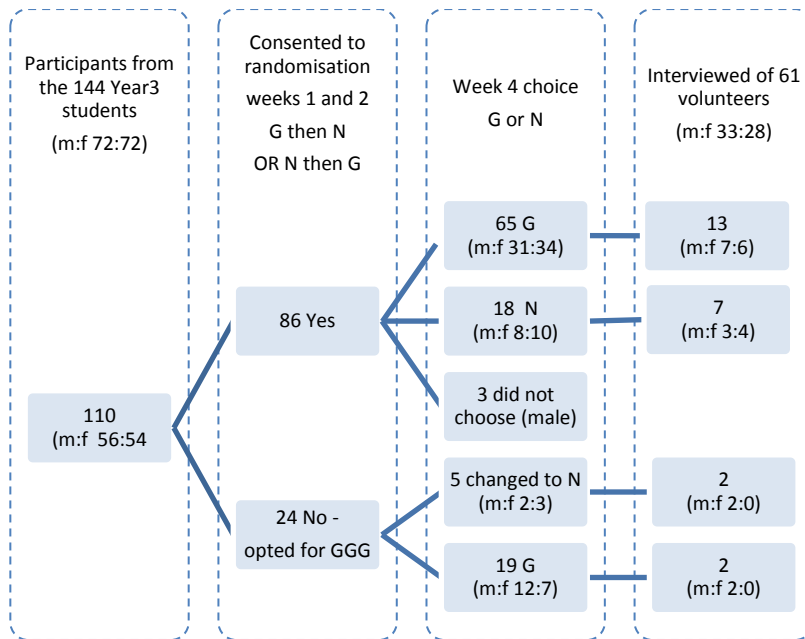
Ethical Approval

The study received approval from the Keele University School of Medicine Ethics Committee.

Results

There were 144 students in the year cohort. 110 (76%) volunteered to participate. 24 of these declined randomisation, opting to have all WPBAs with grades but five of the 24 changed their minds in week 4 and had their final WPBA without grades. 86 students were randomised and 83 of these chose whether or not to have grades for their third and final WPBA; three did not complete the choice questionnaire. Of the 83, 65 (78%) chose to have feedback with grades and 18 (22%) without (Fig 1).

Fig 1 Flow diagram of the participants' passage through the study.
Key: G=WPBA with grades; N=WPBA with no grades, GGG= all three WPBA with grades

Participants from the 144 Year3 students (m:f 72:72)

Consented to randomisation weeks 1 and 2 G then N OR N then G

Week 4 choice G or N

Interviewed of 61 volunteers (m:f 33:28)

110 (m:f 56:54

86 Yes

24 No - opted for GGG

65 G (m:f 31:34)

18 N (m:f 8:10)

3 did not choose (male)

5 changed to N (m:f 2:3)

19 G (m:f 12:7)

13 (m:f 7:6)

7 (m:f 3:4)

2 (m:f 2:0)

2 (m:f 2:0)

The student characteristics of gender, order of randomisation and what grades they got in their initial graded WPBA were facets of the internal context at the point of choosing grades or no grades in the final WPBA. However, these did not appear to influence the choice significantly (Table 1). Recent OSCE results did show an effect on choice but only around the pass/fail divide - all the students who failed the OSCE chose grades in week four, whereas those who barely passed were a little more likely to choose no grades. Numbers in these two sub-groups were too small for analysis of significance but mirrored studies of feedback-seeking behaviour following summative assessment (30,31).

Table 1: Characteristics of students in the randomised cohort making a choice of grades or no grades their 3rd WPBA.

Characteristic		Choice	
		Grades	**No Grades**
Cohort	**All**	65 (78%)	18 (22%)
Gender	**Male**	31 (79%)	8 (21%)
	Female	34 (77%)	10 (23%)
Order of randomisation of WPBA	**1) Grades** **2) No Grades**	30 (73%)	11 (27%)
	1) No Grades **2) Grades**	35 (83%)	7 (17%)
Received Borderline (B) or Must Improve (MI) grades in any domains in the graded WPBA in week 1 or 2	**No B or MI grades**	15 (79%)	4 (21%)
	Did get B or MI	49 (77%)	15 (23%)
Recent OSCE results (if disclosed)	**Clear pass**	50 (79%)	13 (21%)
	Just passed	4 (57%)	3 (43%)
	Failed	10 (100%)	0 (0%)

Sixty-one students volunteered for interview. Of these, 24 were purposively sampled for interview (15 who chose grades and 9 who chose no grades; 14 male and 10 female; 19 passed the recent OSCE, three failed and were preparing to re-sit the OSCE, and two did not disclose their OSCE results for this study).

54 (83%) of 65 students who chose grades and 18 (78%) of 23 who chose no grades answered the question "Why did you choose grades/no grades?"on the choice questionnaire. These responses were included verbatim in the text for analysis and linked to the student's interview text.

Overview of how our initial programme theory changed as a result of the realist evaluation of interviews and questionnaire text responses

Our respondents validated the expected two types of mechanisms resulting from feedback: cognitive mechanisms which were generally associated with desired gains as described by Hattie and Timperley [1]; and motivating and demotivating mechanisms affecting outcomes of effort and self-efficacy, which could conflict as students appraised their feedback. Students differed in which of these mechanisms were activated, and to what extent, by the inclusion of grades in their feedback. In seeking to understand these differences in outcomes such as effort and self-efficacy we found evidence of both selective recall of feedback and selective focusing on feedback. In our resulting programme theory (fig 2), we have likened this selective recall and selective enhancement to the "filtering" and "pumping" mechanisms of a semi-permeable cell membrane which can process feedback as it allows it in. The metaphor of the semi-permeable membrane has been adopted because it is a good representation of what we have found, and will be understood by clinical educators and by medical students. We will postulate that this is how we all receive feedback, the settings of this semi-permeable barrier being attributable to various aspects of the internal and external context including the influence of being graded.

These findings are expanded and discussed below. Quotes from transcripts are identified by M=male F=female; participant ID in the order interviewed, and annotated with their choice of grades or no grades. Unless annotated as a questionnaire comment, the quotes are from interviews.

Findings: Cognitive and motivational mechanisms (representative samples provided in table 2)

Cognitive mechanisms triggered by receiving grades were locating self, clarifying goals and focusing attention. Students who chose grades perceived them as clear additional information regarding their position and progress (quotes 1&3) but might need the context of the same assessor for this to be worth the risk of harm (quote 2) and needed an explanation for the grade if it were low (quote 4). Some students described the importance of an external opinion on position and areas needing attention (e.g. quote 5). The outcome of this clarity was focused identification of strategies for improvement of consultation skills and also training in self-assessment by exit standard (e.g. quote 3).

The counter argument that grades create a focus on the grade rather than on the content of feedback was made by only one student (F23 in quote 6).

Motivational effects of grades were more mixed. Grades could galvanize or reduce effort and this depended on the internal and external context as well as on the grades (quotes 7&8). Outcomes such as enhanced self-efficacy for consultation tasks and acquisition of an identity as a proto-professional were desired by many, and grades had the potential to promote or detract from these. M1 (in quotes 5, 8 &12) describes two motivational effects: grades galvanised effort but diminished self-efficacy at least initially. He was one of the (few) students who had an uncomfortable feedback relationship, describing his GP as "brutally honest" and in that context experienced some loss in self-confidence from the verbal feedback as well as from the low grades. Many students however felt they were in a very supportive relationship. Those who felt they were improving in the context of a supportive relationship described choosing to have grades in the confidence that they would get external validation which would boost their self-efficacy (e.g. quote 11).

To others, grades represented a clear risk of harm. This risk/benefit dilemma triggered varying mechanisms in these Year 3 students when faced with grading by Year 5 exit standards. Many described how being graded to exit standard orientated learning and motivated working to goals, but risked triggering demotivation. The risks of grades causing demotivation were much more apparent for students who chose not to have grades with few describing any compensating cognitive or motivational mechanisms. In particular, grades did not galvanise effort for them, a response which some may have developed because of previous experiences. M22 describes this but was aware of his internal risk/benefit conflict and decided: "I guess grades are a necessary evil" (quote 9). F21 (quote 10) also felt that although grades could be beneficial (if explained) they were too risky: because she trusts her self-assessment, grading could be harmful if dissonant.

Some might have preferred grades in another context but chose no grades in the context of a lack of trust in the assessor or in the assessment system (e.g. quote 13). Others described grades as merely useless and irritating rather than potentially harmful.

Table 2: Inside the 'black box' - what works for whom and why? Context-Mechanism-Outcome (CMO) relationships described by students talking about the effects of grades

CMO configurations	Quote
Giving grades (C): an intervention here seen to be helping students to improve consultation skills (O) by cognitive mechanisms (locating self, clarifying progress, focusing attention) (M)	
Grades (C) help students to locate themselves clearly(M) so they know what needs improving (M), in the context of a trusted and consistent assessor (C)	**Quote 1:** "Although I have found the comments… very helpful, as it provides me with specific points to learn from, I feel the grades are useful as they give a more concrete and measurable representation of my performance." **M12** questionnaire response (chose grades) **Quote 2:** "I've had a change of tutor since the first and second weeks and I feel like there would be a lack of consistency which would be reflected in my grades so I would rather in this case not have grades than have a grade which isn't necessarily reflective of how I may have improved. But if I had had the same tutor, I would have chosen to have grades." **F19** questionnaire response (chose no grades)
Grades (C) help to clarify progress towards goals (M) but need explanation (C) to be useful	**Quote 3:** "whilst I feel feedback is all well and good and it's something that you need as well to sort of take on board and improve, I feel the only way you can see if you <u>have</u> actually taken that feedback on is if you have a quantitative benchmark to work from. ….' So now I have goals for the next two years to work towards." **M5** (chose grades) **Quote 4:** "Everyone has to reach proficiency so I wouldn't mind having grades, but I don't find grades useful unless there is detailed feedback. I can see how they are useful in terms of knowing where you are with respect to the exit… without feedback I can't use the grade. It merely demotivates me. I don't expect to be perfect but I need the feedback to explain the bad grade." **F21** (chose no grades)
Grades (C) focusing attention (M) resulting in increased effort in those target areas (O)	**Quote 5: Interviewer**: "In what way did receiving those grades affect your progress?" "Well then I knew… I had opening, history taking where I had specific areas of the consultation… I'd have to build upon rather than just getting generalised feedback … It sort of highlighted areas I thought were fine. Like I thought my history taking was fine… but the GP said no, you're missing all these things and then I tried to build on those things and make mine a lot more detailed." **M1** (chose grades)
Grades (C) focus attention (M) unhelpfully on the grade in the internal	**Quote 6:** "I just don't like grades at all. I dunno. It's all the competition with other people as well – what did you get? How did you get on? I just find it really annoying. I just find myself more anxious and worried if I've got grades rather than just the no grades. …Because if you put

context of previous experience (C)	grades on something I have a tendency to focus on the grade rather than… anything that was said… So like you could have said ten things that I improved on or did well in, but if you said to me 'But despite that your grade was this' and the grade wasn't necessarily very good I would go 'Oh'. And I would think of it as not very good rather than thinking about the fact that I did a lot of things well." **F23** (chose no grades)

Grades(C): an intervention increasing or decreasing effort (O) by motivational mechanisms (M)

Galvanizing effort (O) by challenging complacency (M) in the context of competitiveness (internal C) or a valued student-assessor relationship (external C)	**Quote 7. "**there were quite a few that were slightly lower than I expected .., it gave me a bit of an idea of where I particularly needed to improve and have a think about which was quite useful and also it gave me more incentive to improve in those areas because I'm horribly competitive." **M4 (chose grades)**
	Quote 8. "I think you're trying to sort of get approval from the person that's assessing you and whatever and make them, you want to know that they think you're sort of improving." **M1** (chose grades)
Grades diminishing effort (O) by complacency (M) in the context of a satisfactory grade (C) or by demoralization (M) with a low grade (C), in a student damaged by previous assessments (internal C).	**Quote 9.** "A grade can be interpreted in an unhelpful way (for me at least). If I do well I won't have the drive to improve my skills further because I'm already up to standard. Likewise If I'm graded badly rather than seeing it as a reason to try harder, I'll be demoralised and unwilling to try the skill at all. This has happened twice already, it took me a while to get over it and start trying the skill again." **M22** questionnaire response (chose no grades)
But with potential to galvanise effort (O) in a different context (C)	**Quote 10.** "I get very nervous about grades and work myself up about them. Bad grades would knock my confidence… Grades can help you rest on your laurels. On the other hand, grades can motivate you to move into the next grade… Maybe the third WPBA in year 4 we should have grades so that if there is a significant need to improve we would have time to do that.." **F21** (chose no grades)

Grades (C) an intervention influencing self-efficacy (O) by demonstrating the presence or absence of progress or competence (M)

Grades (C) showing progress (M) therefore causing gains in self-	**Quote 11.** "I feel that my consultation skills have been improving over the last 4 weeks and I have benefited from the feedback… provided... I think that having grades in the final week will show the progress I have

efficacy (O)	made since week one and give me the confidence to carry this forward to my future medical practice." **M6** questionnaire response (chose grades)
(Risk of) Diminishing self-efficacy (O) in the context of –comparison with exit standard (C);	**Quote 12.** "I've recognised now that being graded against the standard of everyone else in the year, although it may sort of help my confidence a bit more than being marked at a higher level, it (being marked to exit standard) means that I know where I need to improve across sort of more long term goals than just short term comparisons to everyone else." **M1** (chose grades)
Or in the context of social norms and expectations about grades (C)	**Quote 13.** "It sounds… like, you know, you're Borderline, not good enough to be at medical school… going back to my family and saying "Oh I got a Borderline", that doesn't sound very encouraging." **F19** (chose no grades)

How our Programme Model was changed by Realist Analysis - Findings of selective recall and processing which we have termed "Filtering" and "Pumping", and the influence of grading on this.

A related set of mechanisms triggered by factors in the internal and external context (including getting grades) resulted in feedback being unequally assimilated. We were able to compare students' recall of their feedback with the written feedback summary they were sent a few days after each assessment. This helped us to explore what interviewees did with their feedback and why. Some did not recall much of their feedback or grades and were not clear why, but others explained mechanisms which involved their grades. Others paid special attention to parts of their feedback, both verbal and written, because of the grades they received.

Contrast these two reactions to Borderline grades:

➢ "Obviously Borderline is not something that you want but I know they use the system to assess doctors so I thought well I'm not gonna be at that stage yet so I didn't pay that much attention to it." F15 (chose no grades)

➢ "I feel that if you get a Borderline grade it makes you think "I really need to improve on this area" and for me gives me a push to do it! If you don't get grades, for me it doesn't give me the same push to think I really have to improve in this area." F25 Questionnaire comment (chose grades)

There seemed to be a 'semi-permeable barrier' to assimilation of feedback which we have likened to the cell membrane with open or closed channels set to pump or selectively transmit feedback. In this analogy it is a 'thick membrane' with embedded processing capacity. The 'membrane' effects were

driven by self-protection, dealing with dissonance and focusing attention. Examples of these are given in Table 3.

Table 3: How do students handle their grades?

CMO configurations	Example
Filtering (M) for self-protection from risk (O) in the context of low grades by exit standard (C); grades could be shocking	**Quote 14.** "My GP … kept saying don't be disheartened by this cus like although you're only proficient at it you will be very good very soon cus you're going along the right tracks but you won't be very good until you're a F1… I think when you sit down and think oh actually yes at this level I should only be proficient so it's o.k. that I was put as proficient but I think when you think you've done quite well and then someone says you're two or three out of five and you were expecting to be five out of five sort of thing I think it's quite a shock (laughing) so I probably wouldn't have chosen to have grades if I'd known, if I'd seen it before I probably would have preferred just to have the writing cus I think the writing reflects much better on what I was doing rather than the grading." **F10** (chose grades before seeing her week 2 grades)
And in the context of low self-assessment (C)	**Quote 15. Interviewer**: "Would there be any scenario you can think of where you'd say I won't have grades thank you?" "Where I had done rubbish and… I didn't want to know." **F24** (chose no grades)
Processing of the meaning of grades (M) to resolve dissonance (M) in the context of low grades and a trusted assessor (C). Once processed, the grades then triggered the next mechanism – "pumping" of feedback (M) resulting in seeking strategies for improvement (O). **Grades (C) induced a temporary performance**	**Quote 16. Interviewer**: "Did the grades have any impact on you?" "Yes. I felt a bit bad cus I thought especially in third year, I shouldn't be achieving like this since we've just done OSCEs (laughing), like two or three weeks ago. So I thought well I should improve and then erm… My friend was really, really good erm so I just started like looking, picking up things that he picked up and then looking at the GP's feedback a lot. And then over the next two weeks, I improved… And I had to like obviously respect their views because they've been doing it for a long time so… I thought - Well the criticism is not really like bad criticism, it's good criticism to help me sort of help me learn and move forward and improve my skills… I think with the grades it made me feel like I should <u>do</u> something." **M6** (chose grades) **Quote 17.** "I prefer to have grades so I can see what my abilities are at the moment and what I need to improve on. Grades give me reassurance if they are good or motivate me to work harder if they need improvement."

approach (O) but after some processing (M) they enhanced learning effort (O)	**F14** questionnaire response, then in the same student afterwards at interview: "I was concerned that erm I'd been given Borderline for Examination and Management but the GP explained that that's the level I was at - at the moment - in terms of erm a doctor… so I'm not sure if he meant that I was Borderline for a student or what. But I would obviously like to improve them. … In examination I think I tried to erm volunteer to examine more patients … I just wanted to sort of prove that I did know how to examine. … I think Examination is one of those things that takes a lot of practice anyway… I wouldn't expect it to be much above Borderline at the moment anyway because you know it's something to improve on." **F14** (chose grades)
Grades (C) triggering "Pumping" (M) to enhance learning in areas needing effort (O)	**Quote 18.** "If I just had verbal feedback erm… I just think having that Must Improve will keep me sort of in mind that it is something I will need to continue working on. But… if see a grade it sort of sticks in my mind quite a bit more than someone just talking to me." **M1** (chose grades) **Quote 19.** "She didn't want to give me erm for example an excellent rating because otherwise I would be too complacent with that so she gave me Can Be Improved (sic) for each of those categories." **Interviewer**: "And is that true that you would have been complacent?" "Erm perhaps so, perhaps so. ...It was good that at the end of each week we'd get, erm an e-mail of all our results so I could look back on them and improve on them and the fact that she's ticked those boxes, the box that erm I could improve on to get associated symptoms so the next time I went to surgery I acted on those recommendations." **M3** (opted for all grades, was due to resit OSCE)

Filtering for self-protection

Because of the risk of harm from feedback, especially from grades, it would not be surprising if students selectively filtered their feedback. This was demonstrated by F15 (quoted above in the text) who described ignoring her Borderline grades. She then chose not to have grades in the final assessment, rationalising that there was no concern or the tutor would have told her so. She and several other students indicated that they would in future be likely to choose grades now that they understood the exit standard grading system and could therefore make meaning out of it. Others found less than top marks too unpalatable, for example, F10 (quote14) chose grades to gauge where she was, how she could improve and by how much she had improved. She recalled some of her written feedback but could not recall her grades except she remembered not liking them. She

described how her tutor had tried to scaffold her understanding of grading by exit standard but she was unable to make use of the grades in the way she had intended.

Filtering out the grades but making meaning from the feedback seems harmless. What is of more concern is when the filter blocks the entire set of feedback because of risk of harm from the grades. F24 (quote 15) could recall feedback from week 1 and week 4 when she had no grades but all she could recall of week 2 was one Borderline grade. We were able to examine the written feedback which was equivalent on all three weeks excepting the presence or absence of grades. F24 seems to have reacted to this grade by filtering out the entire set of feedback. F23 (quote 6) also indicated that she only saw the grade and not the feedback.

Processing to resolve dissonance

Several students described how they had struggled to make sense of their grades when they did not initially agree with them. M6 and F14 illustrate this (quotes 16&17). M6 in the context of the GP being a trusted advisor had to resolve the dissonance. He described an initial drop in self-efficacy followed by a galvanizing effect of grades with Borderline grades forcing attention on why he was not as good as his peer. F14 wished to prove that she was better than the Borderline grade she had been awarded. She later returned to her initial learning goals as stated in her questionnaire response.

Pumping to enhance learning

There was also evidence of selective "pumping" of the feedback, and students described using their grades to trigger the pump, by focusing attention and effort as illustrated by quotes 18&19. This is contrary to the expectation that grades would prevent students from taking notice of their other feedback (1).

Programme model

So what works for whom and why? The mechanisms induced by grades are complex. Here we have gained a few glimpses into the "black box". During our analysis we refined our programme model (Fig 2). We considered that the act of choosing grades or no grades was an interim outcome which fed back into the context as shown.Not shown in Fig 2 are the other elements of the external context independent of WPBA which were also mentioned as contributing to the learning outcomes such as the types of consultations students were conducting and their feedback from peers at the weekly video feedback session.

Fig 2: Programme model depicting some of the contexts, mechanisms and outcomes found in this study.

See Appendix

Discussion

This is the first study directly comparing feedback from WPBA with and without grades in undergraduate medical education. We found that 78% of our mid-stage medical students want grades and that grades can be helpful when linked to formative assessment. Grades locate the student and calibrate their self-assessment, promoting self-awareness of competence and limits and helping them track progress, especially if referenced to an understood standard such as that expected of the graduating doctor (15). In students who have well-developed self-regulation, low grades can focus attention and effort; satisfactory and high grades can enhance self-efficacy; and in the context of constructive feedback from a trusted assessor, students will usually create constructive meaning out of low grades.

However, we also found that 22% of students preferred not to have grades for a number of reasons including avoiding harm. This suggests that for such students, if grades are given, care must be taken to adjust the external context to diminish the risk of harm and to help the student orient their internal context (self-regulate). In other words, grades are concrete, powerful and blunt and need explanation to help students create helpful meaning from them. Grades can also create complacency: it is easy to focus on the grade and to ignore carefully prepared advice on how to improve. When the student has a prevention approach to learning (22), low grades risk reducing self-efficacy in some; and low grades may encourage others to focus on proving what they can do rather than on looking for ways to improve what they do.

The self-protective filtering, processing and "pumping" we found are comparable with the feedback-seeking behavior found in Veterinary students by Bok *et al* (41). Bok *et al* found that students adjusted their feedback-seeking behavior to avoid harm and promote gain, and the contextual factors promoting or deterring feedback-seeking were the learning climate and relationship with assessors. So we suggest that the "settings" of the protective filters depend on prior experiences and promotion/prevention focus but are influenced by the feedback relationship. The clarity and simplicity of grades make them more difficult to ignore than undesired narrative feedback. Once allowed through the filter, sense must be made of both narrative feedback and grades. This meaning-making is also influenced by gain-seeking and loss-avoidance; hence the students' insistence that a grade must be explained and advice given on how to improve, so that they can turn a potential loss into an educational gain. Students then seem to switch the "pump" on and suck in the constructive criticism attached to the low grade. This again aligns with self-regulation theory - the student's motivational

regulatory focus affects the way they handle feedback. This could spiral in either direction as the feedback affects the motivational focus (40).

Comparing these findings with our initial expectations from the literature, we did find as expected that the most prevalent contextual factors in the positive or negative influence of grading were the perception of tutor support and the student's prevention or promotion approach to and past experiences of assessment (5,22). The finding that potential negative effects were diminished and positive effects were enhanced by selective filtering and pumping modifies the expectation from the literature that grades will neutralise feedback about the task (1,10). While we did find that neutralising of feedback was triggered by grades in a few, and other students mentioned it as a potential adverse mechanism, these were capable of making adjustments to acquire the feedback they felt they needed. This is also a form of self-regulation.

While we acknowledge that other mechanisms could be operational, we postulate that the metaphor of the semi-permeable membrane for the way we handle feedback may be valid and may have resonance for many in medical education.

Strengths and weaknesses of this study

The crossover design has enabled us to examine a complex real-life phenomenon – the effect of including grades in formative feedback, and explore how students both make choices about receiving grades and make sense of feedback which does and does not include grades. The study design enabled sophisticated purposive sampling of students who received feedback in different contexts but could not control for all complexities in the context. There are undoubtedly other influences, but this design has reduced them. While we do not have direct access to the feedback discussions between students and tutors, we did have the written summaries and thus insight as to what was discussed to compare with students' recollections. The number of participants has been sufficient to enable an understanding of the process in the various learner contexts. The approach was obviously acceptable as so many participated.

The study is limited by being in a specific context: a post assessment end of year clinical placement with assessment by exit standard two years before exit with students from one medical school in (usually) supportive longitudinal relationships with tutors. Nevertheless these are also strengths: students should have been motivated by learning as they were not approaching summative assessment, a large proportion received low grades in at least one assessed domain and we were able to analyse their response to this and found that these protective mechanisms are evident even in supportive relationships. Finally, the outcomes reported are self-reported by students but such self-report is a feature of qualitative inquiry.

Implications of the study

Given the prevalence of grades in medical education, this study has wide implications, assuming the findings are replicated in other settings, for example when students are approaching summative assessment.

The conclusions of this study for "best practice" formative WPBA are as follows: Firstly, grades are important to many medical students for a number of reasons. These reasons can be explored by a supportive tutor who can encourage a learning approach to WPBA, aiming for self-awareness of competency and prioritisation of areas for improvement. If it seems that receiving grades will enhance the seeking of strategies for improvement, they should be offered as an element of formative feedback. Secondly, the criteria for allocation of grades must be understood by students who receive them. Thirdly, not all students will accept the offer of grades where choice is available, and this decision should be respected but perhaps explored by tutors. Finally, feedback is processed in the light of previous feedback so all tutors must understand that today's feedback affects the response to tomorrow's feedback. This study provides further evidence of the complexity going on inside the heads of feedback recipients, and helps the push towards a more personalised approach to feedback. We now need to determine whether attention to these factors does indeed help students to make more use of their feedback and enhances learning.

1School of Medicine, Keele University, Keele, UK
2School of Medicine, University of Aberdeen, Aberdeen, UK
Correspondence: Janet Lefroy, School of Medicine, Keele University, Keele, Newcastle under Lyme ST5 5BG, UK.
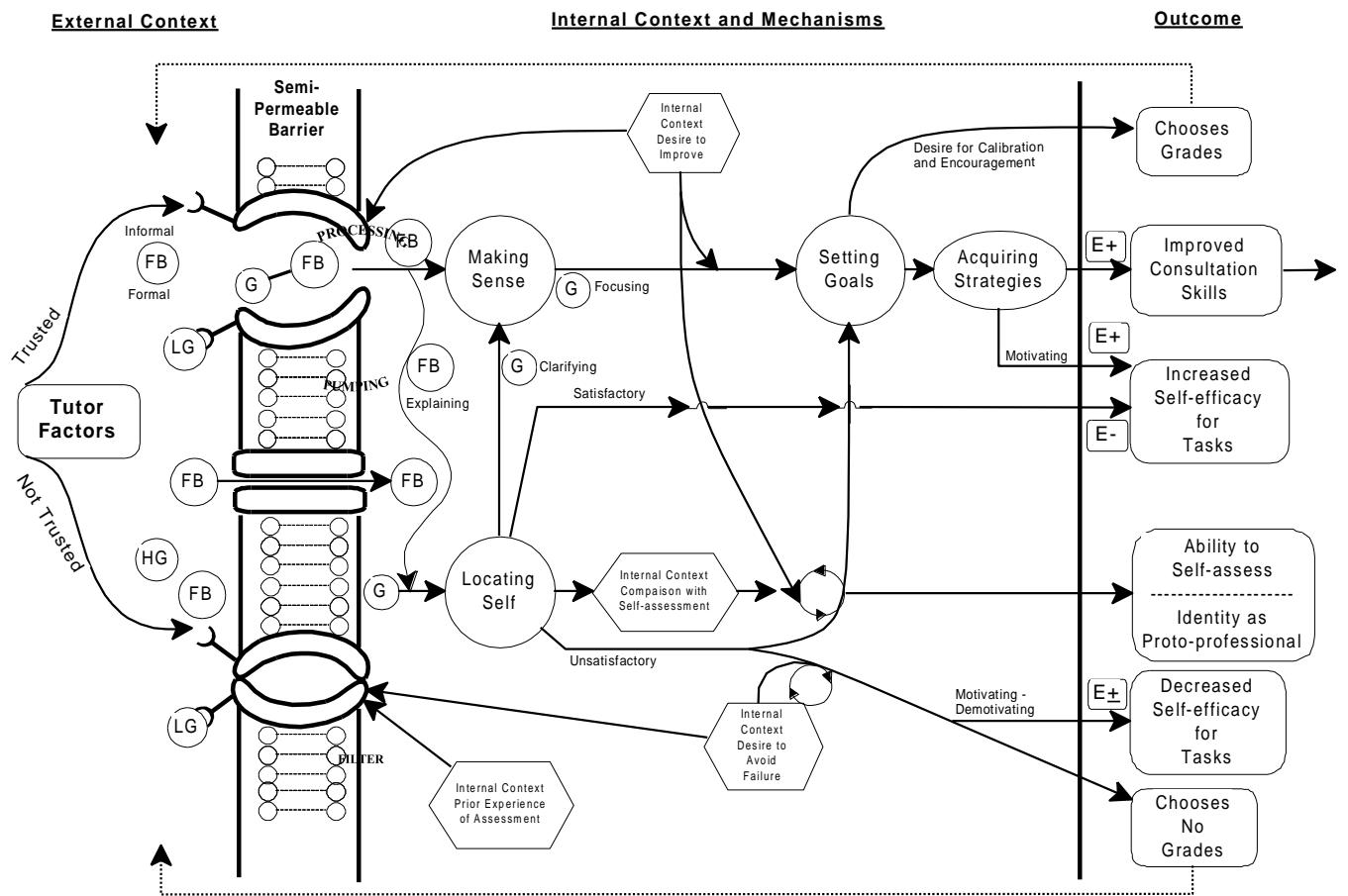Tel: 00 44 1782 734681; E-mail: j.e.lefroy@keele.ac.uk

References

1.   Hattie J, Timperley H. The Power of Feedback. Rev Educ Res. 2007 Mar 1;77(1):81–112.

2.   Norcini J. The power of feedback. Med Educ. 2010;44:16–7.

3.   Carless D, Salter D, Yang M, Lam J. Developing sustainable feedback practices. Stud High Educ. 2011 Jun;36(4):395–407.

4.   Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach. 2007;29(9-10):855.

5.   Veloski J, Boex JR, Grasberger J, 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.

6.   Gibbs G. How assessment frames student learning. In: Bryan, C and Clegg K, editor. Innovative assessment in higher education. London: Routledge; 2006. p. 23–36.

7.   Weaver MR. Do students value feedback? Student perceptions of tutors' written responses. Assess Eval High Educ. 2006 Jun;31(3):379–94.

8.   Higgins R, Hartley P, Skelton A. Getting the Message Across: The problem of communicating assessment feedback. Teach High Educ. 2001;6(2):269.

9.   Poulos A, Mahony MJ. Effectiveness of feedback: the students' perspective. Assess Eval High Educ. 2008 Apr;33(2):143–54.

10.   Kluger A, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. . Psychol Bull. 1996;119(2):254.

11.   Shute VJ. Focus on Formative Feedback. Rev Educ Res. 2008 Mar 1;78(1):153–89.

12.   Urquhart LM, Rees CE, Ker JS. Making sense of feedback experiences: a multi-school study of medical students' narratives. Med Educ. 2014 Feb;48(2):189–203.

13.   Norcini JJ, Blank LL, Duffy FD, Fortna GS. The Mini-CEX : A Method for Assessing Clinical Skills. Ann Intern Med. 2003;138(6):476–81.

14.   McKinley RK, Fraser RC, Van V der, Hastings AM. Formative assessment of the consultation performance of medical students in the setting of general practice using a modified version of the leicester assessment package. Med Educ. Department of General Practice and Primary Health Care, University of Leicester, Leicester General Hospital, Gwendolen Road, Leicester LE5 4PW, UK; PM:10886641; 2000;34(7):573–9.

15.   GMC. Tomorrow ' s Doctors The duties of a doctor registered with the General Medical Council. 2009.

16.   Eva K, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med. 2005;80:46.

17.   Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. Adv Health Sci Educ Theory Pract. 2012 Mar;17(1):15–26.

18.   Murdoch-Eaton D, Sargeant J. Maturational differences in undergraduate medical students' perceptions about feedback. Med Educ. 2012 Jul;46(7):711–21.

19.   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. Med Educ. 2012 Jun;46(6):593–603.

20.   Molloy E, Boud D. Seeking a different angle on feedback in clinical education: the learner as seeker, judge and user of performance information. Med Educ. 2013 Mar;47(3):227–9.

21.   Sitzmann T, Ely K. A Meta-Analysis of Self-Regulated Learning in Work-Related Training and Educational Attainment: What We Know and Where We Need to Go. Psychol Bull. 2011;137(3):421.

22.   Kluger A, Van Dijk A. Feedback, the various tasks of the doctor, and the feedforward alternative. Med Educ. 2010;44:1166.

23.   Govaerts M, Van der Vleuten C, Schuwirth L, Muijtjens A. Broadening Perspectives on Clinical Performance Assessment: Rethinking the Nature of In-training Assessment. Adv Heal Sci Educ. 2007;12:239.

24.   Monrouxe L, Rees C, Lewis N, Cleland J. Medical educators' social acts of explaining passing underperformance in students: a qualitative study. Adv Heal Sci Educ. 2011;16(2):239.

25.   Cleland JA, Knight L V, Rees CE, Tracey S, Bond CM. Is it me or is it them? Factors that influence the passing of underperforming students. Med Educ. 2008 Aug;42(8):800–9.

26.   Holmboe E, Yepes M, Williams F, Huot S. Feedback and the mini clinical evaluation exercise. J Gen Inter Med. 2004;19:558.

27.   Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently: mechanisms that contribute to assessor differences in directly-observed performance assessments. Adv Health Sci Educ Theory Pract. 2013 Aug;18(3):325–41.

28.   Altahawi F, Sisk B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: Experience in a competency-based portfolio system. Med Teach. 2012;34:221–5.

29.   Manzone J, Tremblay L, You-Ten KE, Desai D, Brydges R. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. Med Educ. 2014 Apr 9;48(4):430–40.

30.   Harrison CJ, Könings KD, Molyneux A, Schuwirth LWT, Wass V, van der Vleuten CPM. Web-based feedback after summative assessment: how do students engage? Med Educ. 2013 Jul;47(7):734–44.

31.   Sinclair HK, Cleland JA. Undergraduate medical students: who seeks formative feedback? Med Educ. 2007 Jun;41(6):580–2.

32.   Higgins E. Making a good decision: Value from fit. Am Psychol. 2000;55(11):1215.

33.   Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011 Jan;33(6):478–85.

34.   Galbraith RM. Got feedback? Med Educ. 2013 Mar;47(3):224–5.

35.   Pawson R, Tilley N. Realistic Evaluation. London: Sage; 1997.

36.   Wong G, Greenhalgh T, Westhorp G, Pawson R. Realist methods in medical education research: what are they and what can they contribute? Med Educ. 2012 Jan;46(1):89–96.

37.   Lefroy J, Gay SP, Gibson S, Williams S, McKinley RK. Development and face validation of an instrument to assess and improve clinical consultation skills. Int J Clin Ski. 2011;5(2):115–25.

38.   Lefroy J, Thomas A, Harrison C, Williams S, O'Mahony F, Gay S, et al. Development and face validation of strategies for improving consultation skills. Adv Health Sci Educ Theory Pract. 2014 Jan 22;

39.   Pawson R. Evidence-based policy: a realist perspective. London: Sage; 2006.

40. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006 Apr;31(2):199–218.

41. Bok HGJ, Teunissen PW, Spruijt A, Fokkema JPI, van Beukelen P, Jaarsma D a DC, et al. Clarifying students' feedback-seeking behaviour in clinical clerkships. Med Educ. 2013 Mar;47(3):282–91.

42. Lefroy J, Hawarden A, Gay SP, McKinley RK, Cleland J. Grades in formative workplace-based assessment: a study of what works for whom and why. Med Educ. 2015;49:307–20.

Appendix : Fig 2

Semi-Permeable Barrier

PROCESSING

Informal

FB

Formal

G   FB   FB

LG

Trusted

Not Trusted

Tutor Factors

PUMPING

FB   Explaining

FB   FB

HG

FB

G

LG

FILTER

Internal Context Desire to Improve

Making Sense

G   Focusing

G   Clarifying

Satisfactory

Locating Self

Internal Context Compaison with Self-assessment

Unsatisfactory

Internal Context Prior Experience of Assessment

Internal Context Desire to Avoid Failure

Desire for Calibration and Encouragement

Setting Goals

Acquiring Strategies

Motivating

E+

E+

E-

Motivating - Demotivating

E+

Chooses Grades

Improved Consultation Skills

Increased Self-efficacy for Tasks

Ability to Self-assess

Identity as Proto-professional

Decreased Self-efficacy for Tasks

Chooses No Grades

Context - Mechanism - Outcome configurations illustrated

Initial Programme Theories confirmed:
1. Feedback has sense made of it leading to goal-setting for areas needing improving. This leads to acquiring strategies for improvement and the final outcome is improved consultation skills.
2. Acquiring strategies for improvement and noticing improvement also increases self-efficacy.
3. Grades enable the student to locate him/herself against the goals thus clarifying the meaning made of the descriptive feedback.
4. Satisfactory grades increase self-efficacy and can cause bypassing of the goal-setting route.
5. Unsatisfactory grades can also have negative outcomes by demotivating.

Additional CMO configurations found:
6. Feedback channels in the semi-permeable barrier may be opened or closed by the degree of trust in the tutor.
7. Feedback channels may also be opened or closed in response to low grades, as set by the internal context – desire to improve or desire to avoid failure, and prior experience of assessment.
8. Grades need explanation so draw attention to relevant feedback i.e. cause "pumping" of feedback.
9. Grades focus goal-setting.
10. Unsatisfactory grades trigger different mechanisms depending on the internal context – in the context of a desire to improve, low grades enhance goal –setting; in the context of desire to avoid failure, low grades decrease self-efficacy.
11. The initial motivating effect of low grades in the performance goal approach turns to demotivation if the additional effort to perform doesn't result in higher grades. This contrasts with the learning goal approach where low grades motivate additional effort to learn.
12. The interim outcome of choice of grades results from desire for calibration and/or encouragement triggered by previous grades.
13. The interim outcome of choice not to have grades results from either the desire to avoid harm, or inability to make use of the grades for a number of reasons (not shown to avoid over-complicating the diagram).