

Feedback in the context of high-stakes assessment

Citation for published version (APA):

Harrison, C. J. (2017). *Feedback in the context of high-stakes assessment: can summative be formative?* [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20170920ch>

Document status and date:

Published: 01/01/2017

DOI:

[10.26481/dis.20170920ch](https://doi.org/10.26481/dis.20170920ch)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

FEEDBACK IN THE CONTEXT OF HIGH-STAKES ASSESSMENT:

CAN SUMMATIVE BE FORMATIVE?

The research reported here was carried out at:



in:



© copyright Christopher Harrison, Maastricht 2017

Printing: Datawyse | Universitaire Pers Maastricht

ISBN 978 94 6159 735 9

FEEDBACK
IN THE CONTEXT OF HIGH-STAKES ASSESSMENT:
CAN SUMMATIVE BE FORMATIVE?

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday 20th September 2017, at 12.00 hours

by

Christopher Harrison

Supervisors

Prof. dr. C.P.M. van der Vleuten

Prof. dr. L.W.T. Schuwirth

Prof. dr. V.J. Wass, Keele University UK

Co-supervisor

Dr. K.D. Könings

Assessment committee

Prof.dr. S. Heeneman (chair)

Dr. G.J. Bok, Faculty of Veterinary Medicine, University of Utrecht

Prof.dr. E.W. Driessen

Dr. M.J.B. Govaerts

Prof.dr. A.D.C. Jaarsma, UMC Groningen

*For my wife, Ilona, and our daughter Scarlett-Maria:
your love, patience and support made this possible.*

Contents

1	Introduction	9
2	Web-based feedback after summative assessment: how do students engage? <i>Published in Medical Education 2013; 47: 734-744</i>	19
3	Barriers to the uptake and use of feedback in the context of summative assessment <i>Published in Advances in Health Science Education 2015; 20: 229-245</i>	43
4	Factors influencing students' receptivity to formative feedback emerging from different assessment cultures <i>Published in Perspectives on Medical Education 2016; 5: 276-284</i>	65
5	Changing the culture of assessment: the dominance of the summative assessment paradigm <i>Published in BMC Medical Education 2017; 17: 73</i>	81
6	Discussion	105
	Summary	121
	Samenvatting	129
	Valorization	137
	Acknowledgements	143
	Curriculum Vitae	147

Chapter 1

Introduction

The central theme of this thesis concerns the challenges of receiving feedback within the context of summative assessment. It is a common assumption that assessments have *either* a formative *or* a summative purpose and that it is difficult for a single assessment to serve both purposes simultaneously.¹ However, is it necessary to maintain the dichotomy between formative and summative assessment, or, to put it more positively, would there be benefits in combining them? To phrase it more succinctly, can summative assessments be formative?

For clarity, it may be helpful to define what is meant by summative and formative in this context. Summative assessments are designed to evaluate whether students have learnt the intended learning outcomes of a course. The distinguishing feature of summative assessments is that the results are used to make decisions about progression to the next stage of a course, or to gain a qualification or satisfy an external regulator. For this reason they are often referred to as high-stakes assessments. Formative assessments are designed to help students monitor their own progress and support their further learning; they do this via the process of feedback. Feedback is typically 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.*²

Feedback can be an extremely powerful phenomenon,³ especially when it is positively-focussed and oriented towards a task rather than an individual.⁴ However, feedback often fails to reach its potential and it does not always lead to performance improvements.⁵ This may be because the feedback is delivered poorly, it may be due to problems with the reception of feedback, or because of a learner's failure to translate the feedback into actual learning activities that bring about a desired change. The coaching relationship between a supervisor and a learner is also increasingly seen as significant.⁶

What would the potential advantages be in combining formative and summative in one assessment?

There are a number of reasons why it may make sense to combine the formative and summative purposes of assessment. Summative assessment plays a pivotal role in determining competence to enter the medical profession. Much effort has been deployed in order to develop psychometrically defensible assessment systems with the aim of determining which students are competent to graduate.¹ Standard-setting procedures have been developed and refined so as to define and recognise the minimally competent student.⁷ These procedures usually allow a student to graduate if he has achieved an overall grade of competence.

High-stakes assessments typically create large amounts of data regarding a student's strengths and weaknesses. Two students with the same overall mark will typically have very different profiles of strengths and weaknesses. Furthermore, this overall mark may well disguise significant areas in which the student is not yet competent. In the binary pass-fail nature of summative assessments, a student who is only minimally competent is treated the same as one who is maximally competent: both can progress to the next stage of learning or graduate as a doctor. There is, therefore, a risk of providing false reassurance about a candidate's ability, both to the student but perhaps more importantly to patients; candidates who pass the assessment are simply labelled as competent. Although this label may be nuanced, perhaps by the provision of grades, the manner in which it is used is crucial: the learner is either good enough to progress or not. In addition, the implicit message given to learners is that those areas in which they did not excel must be unimportant as they have been labelled as competent overall.⁸ Reducing all the information obtained from an assessment down to a binary pass-fail verdict simplifies high-stakes decision-making. Binary decisions on individual assessments are often then combined to make decisions about overall progression through a course. While this is an understandable approach, it has the potential to mislead as much rich information about a learner has been lost in the process. More importantly, this approach is unlikely to help students prepare for independent self-regulated learning as doctors. As independent professionals, they will need to be able to identify their own strengths and weaknesses; taking responsibility to analyse detailed feedback should help prepare them for this, but simple knowledge of pass-fail decisions will provide little assistance.⁴ There is also a financial consideration: practical assessments in medical education are particularly expensive, so it makes sense to maximise the value which can be obtained from them.⁹

More positively expressed, all the assessment data created represents an opportunity for students to learn from the assessment. The mantra that 'assessment drives learning' is often repeated, with the assumption that the effect is always beneficial. However, Newble and Jaeger¹⁰ provided empirical evidence that examinations can sometimes encourage unhelpful learning behaviours in undergraduate medical education. Dochy and colleagues¹¹ have considered the link between assessment and learning from a constructivist viewpoint and demonstrated that the relationship is complex. They described pre-assessment, pure assessment and post-assessment effects on learning. Pre-assessment effects refer to the learning which takes place before an assessment. Pure assessment effects refer to the learning that actually occurs during assessments, such as the compilation of portfolios or test-enhanced learning.¹² Post-assessment effects refer to the learning which takes place as a result of feedback from an assessment; these are what this thesis is principally concerned about.

Why might it be difficult to combine formative and summative elements in one assessment?

There are, however, potential difficulties in combining formative and summative elements into one assessment. Firstly, there are significant practical difficulties with delivering feedback in this context. High-stakes assessments typically involve large numbers of learners sitting an assessment at the same time. The large amount of data which are generated from an assessment needs to be processed. The urgent need for an institution is to determine progression decisions resulting from the assessment; feedback information is likely to be seen as of secondary importance. Faculty resource

es are finite and therefore typically focus on candidates who have failed an assessment.¹³ The timing of such assessments is often problematical; high-stakes exams usually occur at the end of a year and students then move to a different part of the course or even leave the institution. There is good evidence that feedback must be given in a timely manner,¹⁴ yet these challenges make this very difficult to achieve. As high-stakes assessments are by definition used for important progression decisions, many institutions are understandably reluctant to divulge their bank of questions to the cohort of learners, for fear that it will spread ‘down the line’ to future students and thus impair the integrity of the assessment. As a result, this can compromise the level of detail which could be provided as feedback.

There are ways to mitigate some of these practical challenges. To address the security of the question bank, many institutions resort to giving grades, even though these have limited influence on subsequent student performance.⁴ Other organisations have circumvented this problem by providing students with details of the learning objectives without revealing the actual questions themselves.¹⁵

Timeliness of feedback delivery can to some extent be addressed by the use of advances in computer technology in order to reduce processing time of assessment data. For instance, the use of tablets, such as iPads, in OSCEs enables the capture of examiners’ verbal feedback about candidates; this can be subsequently downloaded from a website, saving significant faculty resources.¹⁶ Use of multi-media approaches, presenting information in relatively simple formats, is advocated in Shute’s review of feedback and is also in line with theories of cognitive load.^{4, 17}

So, while delivery of feedback after summative assessments appears to be challenging, the problems appear surmountable. The reception of the feedback presents a greater challenge. At the most fundamental level learners may not bother to read feedback that is offered to them. Perhaps surprisingly, receptivity to feedback in the context of summative assessment has rarely been studied. In one study, Sinclair and Cleland¹⁸ found that more than 50% of students did not take up an offer to receive feedback after a high stakes assessment. As a single study, it is unclear if this is a generalizable finding. A limitation of this interesting

research is that it looked purely at whether students picked up the feedback; it was not able to determine what, if anything, students did with the information, though this is clearly of much more importance than simply picking up the feedback. This paper also hinted at a link between learners' performance and their use of feedback.

As there has been a lack of research into feedback receptivity in the context of summative assessment, it may be helpful to look at receptivity to feedback in other contexts for clues to help understand the potential challenges. Individuals' receptivity to feedback can be divided into four broad categories: they may explicitly and directly seek out feedback, they may indirectly seek it by a process of monitoring others' reactions to them, feedback may be given without being actively sought, or they may deliberately avoid seeking feedback. The concept of pro-active feedback seeking has been much studied in the organisational psychology literature, with less focus on medical settings.¹⁹⁻²¹ The predominant focus has been on the individuals concerned, namely the feedback-seeker and the feedback-giver. Learners face a dilemma; they want more feedback, but fear receiving disconfirming information which challenges their own self-assessment.²¹ The emotional impact of receiving feedback is increasingly recognised. Learners can avoid seeking feedback for fear of "looking stupid".²² Critical feedback can unleash powerful emotions which have to be dealt with before the learner is receptive to the actual feedback content.^{22, 23} Although learners do not receive feedback as individuals within a vacuum, there has been less attention on the context in which feedback-seeking may or may not occur. One study reported that individuals were less likely to seek feedback in a public setting with a potentially evaluative audience.²⁴ Transferring this finding to a medical context, Crommelinck and Anseel¹⁹ argue for the use of information technology to support feedback-seeking in a private setting.

The context in which feedback occurs within medical education can perhaps best be understood by comparing its feedback culture with other educational settings in which learners receive feedback. In a series of studies, Watling²⁵⁻²⁷ has compared medicine's learning culture with music education and sport coaching. Medical learners valued feedback to achieve a level of competence and to achieve a state of autonomous practice in which further observations would not be required. By contrast, trainees in music and sport had expectations of receiving more critical feedback and being pushed to an ever-better performance. Medical supervisors of learners often had dual roles of coach and assessor which potentially compromised the learners' receptivity to feedback. Although these studies looked at medicine's feedback and learning cultures in general, they did not consider feedback specifically within a summative assessment context. However, they do indicate that context, and cultural expectations, appear to have significant influences on receptivity to feedback.

Although the post-assessment effects of summative assessment on subsequent learning have been unclear, there are some worrying indications from the profound influences that

summative assessments have on the learning beforehand. Both within medical education and in the wider educational establishment, there is evidence that summative assessments drive learners towards adopting undesirable superficial learning strategies and so-called “binge-learning” shortly before taking tests.²⁸⁻³⁰ Cilliers and colleagues²⁹ argue that the overpowering nature of high-stakes assessments on learning is likely to overwhelm the positive effects on learning that emanate from formative assessment and feedback.

Given the undesirable learning effects from summative assessments, and the inherently flawed reliability of each single assessment, there have been calls for a complete change in assessment culture, away from a concentration on assessment *of* learning and towards assessment *for* learning. A programmatic approach to assessment was first proposed more than 10 years ago.³¹ Since that time, quality criteria and guidelines for such programmes have been developed.³² More recently a theory-driven framework for the practical implementation of a programmatic assessment model has been proposed.³³ This model is designed for learner-centred programmes which favour deep learning strategies and is designed to have three main purposes: it should support assessment for learning; it should enable high-stakes decisions to be made using robust evidence; it should produce information that enables the curriculum to be improved. While this has been successfully implemented in some centres,^{34, 35} its introduction elsewhere has not always been straightforward. Empirical work has demonstrated that students may still regard the low-stakes assessments as summative hurdles to be overcome instead of learning opportunities.^{36, 37} It is therefore unclear how well these modifications to the assessment system promote learning from feedback.

Even if institutions were to contemplate a change in assessment culture, the potential challenges are immense. Both individuals and organisations are typically inherently resistant to change. Key members of organisations usually share common collective beliefs and assumptions about an organisation, often referred to as a paradigm.³⁸ These are often implicit and “taken for granted” and lead any potential changes to be viewed through the prism of the existing organisational paradigm. In addition, individuals have their own beliefs about proposed changes and the potential consequences for their own roles.³⁹ While the practical challenges involved in implementing a problem-based learning curriculum have been studied,⁴⁰ there has as yet been little empirical work about the implementation challenges of considering a change in assessment culture.

In summary, there are reasons why it would appear sensible to use the wealth of data from summative assessments to provide feedback to learners, allowing them a nuanced view of their strengths and weaknesses rather than a simple pass-fail decision. Although the delivery of feedback in these circumstances presents significant practical problems, these can be overcome, in part by the use of information technology. However, learners’ receptivity to the feedback is more challenging; it is uncertain whether they will engage with the feedback in a meaningful way in this context. There are potential barriers at both

the individual and cultural level to the reception of feedback, but these have not been studied in the context of summative assessment. Summative assessment can have adverse effects on other aspects of the learning process, so it is plausible that it could have undesirable effects on feedback receptivity. Alternatives to the summative assessment culture towards one based on an assessment for learning have been proposed, but there is a lack of empirical evidence as to how this can be successfully implemented in a way that engages relevant stakeholders.

Research Questions

There are therefore several overarching research questions which arise from the theoretical framework and which this thesis seeks to address:

1. How do students engage with feedback in a high stakes context?
2. What are the influences and barriers to feedback receptivity, both at an individual and institutional level, in this context?
3. Can we change the assessment culture to one which prioritises and supports learning from feedback? What is stopping us?

Thesis Outline

The study described in Chapter 2 sought to address the first main research question by evaluating the extent to which medical students engaged with feedback after a high stakes clinical assessment. To address the problems of limited faculty resources and the need for timely feedback delivery, feedback information was provided in a web-based format. As well as helping with feedback delivery, the web-based format also enabled students' use of the information to be monitored. This website usage data was linked to students' performance in the assessment and their responses to a questionnaire concerning learning-related student characteristics. As an additional feature, the website provided guidance to learners on formulating an action plan following study of the feedback.

While this quantitative study provided useful information on how students engage with feedback in a summative context, it failed to provide an explanation of the reasons why learners may, or may not, engage with feedback offered in these circumstances. Chapter 3 describes a study using qualitative methods which sought to explore the influences and barriers to feedback receptivity at an individual level after a high stakes assessment. The results of this study highlighted the dominant effect of the summative assessment culture and a disconnection between assessments and subsequent learning.

The study in Chapter 4 therefore looks at the factors within medical schools' assessment cultures which aid or hinder receptivity to feedback. Six focus groups were conducted in three medical schools, in three separate countries, with different institutional approaches to assessment, ranging from a traditional summative assessment structure to a fully implemented programmatic assessment system. This study highlighted a need to change aspects of the assessment culture in order to enhance receptivity to feedback.

Chapter 5 describes a study, using participatory design techniques, in which a mixed group of stakeholders are asked to redesign a summative assessment culture in order to ensure that students would make more use of post-assessment feedback. The challenges involved in considering a change in assessment culture, and the dominance of the summative assessment paradigm, are explored.

Finally, Chapter 6 contains a discussion of the main findings of the studies in the light of the three main research questions, addresses their strengths and limitations and considers the implications for medical education and research.

Context

The setting for much of the research was Keele University Medical School, UK. It is one of the newest British medical schools, graduating its first cohort of doctors in 2012. One of the key aims of the curriculum was to develop a programme that created graduates with academic and clinical excellence, not just competence. Progression through, and exit from, the course relied on a series of high stakes assessments. At the time the research was beginning, the faculty leadership was actively seeking to introduce more feedback after these assessments. This environment was therefore felt to provide a very suitable setting for this research. The need to consider receptivity to feedback in different assessment cultures led to research also taking place in Cleveland Clinic Lerner College of Medicine, USA and the Physician-Clinical Investigator Programme at Maastricht University, Netherlands. These both take a programmatic approach to assessment.

Please note that the chapters were all submitted to journals as separate articles to be read on their own. There is therefore inevitably some repetition across chapters.

References

1. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet* 2001;357:945-9.
2. Van de Ridder JM, Stokking KM, McGaghie WC, Ten Cate OT. What is feedback in clinical education? *Medical Education* 2008;42:189-97.
3. Hattie J, Timperley H. The power of feedback. *Review of Educational Research* 2007;77:81-112.
4. Shute VJ. Focus on formative feedback. *Review of Educational Research* 2008;78:153-89.

5. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;119:254-284.
6. Sargeant J, Lockyer J, Mann K, et al. Facilitated reflective performance feedback: developing an evidence- and theory-based model that builds relationship, explores reactions and content, and coaches for performance change (R2C2). *Academic Medicine* 2015; 90:1698–1706
7. Norcini JJ. Setting standards on educational tests. *Medical Education* 2003;37:464-9.
8. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, Regehr G. Towards a program of assessment for health professionals: from training into practice. *Advances in Health Sciences Education* 2016; 21: 897-913.
9. Brown C, Cleland J, Walsh K. The costs of medical education assessment. *Medical Teacher* 2016;38:111.
10. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Medical Education* 1983;17:165-71.
11. Dochy F, Segers M, Gijbels D, Struyven K. Assessment Engineering: Breaking down barriers between teaching and learning, and assessment. In: Boud D, Falchikov N, (Eds.), *Rethinking assessment in higher education: learning for the longer term*. Oxford: Routledge; 2007. p 87-100.
12. Larsen DP, Butler AC, Roediger III HL. Test-enhanced learning in medical education. *Medical Education* 2008;42:959-66.
13. Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: theoretical and methodological insights from a systematic review. *Medical Education* 2013;47:242-51.
14. Lefroy J, Watling C, Teunissen PW, Brand P. Guidelines: the do's, don'ts and don't knows of feedback for clinical education. *Perspectives on Medical Education* 2015;4:284-99.
15. Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: feedback for individuals and institutions based on the progress test. *Medical Teacher* 2010;32:486-90.
16. Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Medical Teacher* 2015;37:323-6.
17. Sweller J, Van Merriënboer JJ, Paas FG. Cognitive architecture and instructional design. *Educational Psychology Review* 1998;10:251-96.
18. Sinclair HK, Cleland JA. Undergraduate medical students: who seeks formative feedback? *Medical Education* 2007;41:580-2.
19. Crommelinck M, Anseel F. Understanding and encouraging feedback seeking behaviour: a literature review. *Medical Education* 2013;47:232-41.
20. Teunissen PW, Stapel DA, Van der Vleuten C, Scherpbier A, Boor K, Scheele F. Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine* 2009;84:910-7.
21. Mann K, Van der Vleuten C, Eva K, Armson H, Chesluk B, Dorman T, Holmboe E, Lockyer J, Loney E, Sargeant J. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Academic Medicine* 2011;86:1120-7.
22. Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17:15-26.
23. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education* 2008;13:275-88.

24. Ashford SJ, Northcraft GB. Conveying more (or less) than we realize: The role of impression-management in feedback-seeking. *Organizational Behavior and Human Decision Processes*. 1992;53:310-34.
25. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Medical Education* 2013;47:842-50.
26. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education* 2013;47:585-94.
27. Watling C, Driessen E, Van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Medical Education* 2014;48:713-23.
28. Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, Van der Vleuten CP. The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education* 2010;15:695-715.
29. Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, Van der Vleuten CP. A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education* 2012;17:39-53.
30. Cilliers FJ, Schuwirth LW, Van der Vleuten CP. A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Medical Education* 2012;12:1.
31. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Medical Education* 2005;39:309-17.
32. Dijkstra J, Van der Vleuten CP, Schuwirth LW. A new framework for designing programmes of assessment. *Advances in Health Sciences Education* 2010;15:379-93.
33. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, Van Tartwijk J. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34:205-14.
34. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007;82:493-502.
35. Driessen EW, Van Tartwijk J, Govaerts M, Teunissen P, Van der Vleuten CP. The use of programmatic assessment in the clinical workplace: A Maastricht case report. *Medical Teacher* 2012;34:226-31.
36. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, Van Beukelen P, Van der Vleuten CP, Jaarsma DA. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education* 2013;13:1.
37. Heeneman S, Oudkerk Pool A, Schuwirth LW, Van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Medical Education* 2015;49:487-98.
38. Johnson G. Managing strategic change—strategy, culture and action. *Long Range Planning* 1992;25:28-36.
39. Kavanagh MH, Ashkanasy NM. The impact of leadership and change management strategy on organizational culture and individual acceptance of change during a merger. *British Journal of Management* 2006;17:S81-103.
40. Des Marchais JE, Bureau MA, Dumais B, Pigeon G. From traditional to problem-based learning: a case report of complete curriculum reform. *Medical Education* 1992;26:190-9.

Chapter 2

Web-based feedback after summative assessment: how do students engage?

Christopher Harrison
Karen D. Könings
Adrian Molyneux
Lambert Schuwirth
Val Wass
Cees van der Vleuten

Published in:
Medical Education
2013; 47: 734-744

Abstract

Context: There is little research into how to deliver summative assessment student feedback effectively. The main aims of this study were to clarify how students engage with feedback in this context and to explore the roles of learning-related characteristics and previous and current performance.

Method: A website was developed to deliver feedback about the OSCE in various formats: station-by-station or on skills across stations. 138 students (in the third year out of five) completed a questionnaire about goal orientation, motivation, self-efficacy, control of learning beliefs and attitudes to feedback. Individual website usage was analysed over an 8-week period. Latent class analyses were used to identify profiles of students, based on their use of different aspects of the feedback website. Differences in learning-related student characteristics between profiles were assessed using ANOVAs. Individual website usage was related to OSCE performance.

Results: 132 students (95.7%) viewed the website. The number of pages viewed ranged from 2 to 377 (median 102). 50% of students engaged comprehensively with the feedback, 27% used it in a minimal manner while a further 23% used it in a more selective way. Students who were comprehensive users of the website scored higher on the Value of Feedback scale while students who were minimal users scored higher on extrinsic motivation. Higher-performing students viewed significantly more webpages showing comparisons with peers than weaker students did. Students who just passed the assessment made least use of the feedback.

Conclusions: Higher performing students appeared to use the feedback more for positive affirmation than for diagnostic information. Those arguably most in need engaged least. We need to construct feedback after summative assessment in a way that will more effectively engage those students who need the most help.

Introduction

Assessment has complex effects on learning, which have been categorised as pre-assessment, post-assessment and pure assessment effects.¹ As a post-assessment effect, the importance of feedback is clearly established.² Most of the literature refers to the role of feedback in workplace-based assessments, following low-stakes assessments or in preparation for high-stakes tests. The role of feedback after summative or high-stakes assessments has received little attention.³ Instead, the focus has been on determining the pass-fail cut score and on the assessment of competence rather than on encouraging excellence. Medical schools also face practical obstacles to providing feedback in these circumstances. First, summative assessment tends to be timed at the end of the course or module, limiting the value of feedback, but also the cohorts are increasingly large and feedback requires considerable faculty resources, so schools have tended to focus on the small number of students who fail assessments. As a result, students' response to feedback after passing summative assessments is not clearly understood.

If students receive no feedback after summative assessment, but are aware that the primary focus of the assessment is to determine whether or not they have achieved minimal competence, there is an inevitable risk that they will prepare for the assessment by concentrating on the relatively low hurdle of minimally acceptable competence, rather than aspiring to attain the more challenging goal of optimal competence, as the latter is not normally rewarded. Similarly, within medical schools, much effort has been deployed to develop psychometrically defensible summative assessment systems to determine which students are competent to graduate.⁴ Standard-setting procedures, which define the minimally competent student, usually allow a student to graduate if he/she has achieved an overall grade of competence.⁵ However, the lower such an overall mark is, the higher is the probability that there remain significant areas in which the student is incompetent. Feedback after summative assessment would offer the opportunity, or even the obligation, for the student to address these areas of weakness. It is therefore important to understand how to engage students with feedback after summative assessments.

Even if students are provided with feedback, it is by no means certain that it will lead to improvements in performance.⁶ There remains considerable uncertainty about what makes feedback effective for learning. Indeed, even the task of ensuring the student receives the feedback is not straightforward. In some studies, more than 50% of students do not take up offers to receive feedback.⁷ Learners may explicitly ask for feedback or may seek it more indirectly, perhaps by monitoring others' reaction to them and their performance, or comparing their results with others.⁸ Learners express a desire for more feedback but fear receiving information which challenges their own positive self-assessment of their abilities.⁹ Some learners appear to want feedback as a means of building their confidence rather than as a means to correct a knowledge or skill deficiency.¹⁰ It is increasingly recognised that there needs to be a two-way communication process between teachers and

learners and a focus on how feedback is received (feedback as a dialogue).^{10,11} Receiving more negative feedback can have a profound effect on students' self-perception.¹² Arguably, the emotional impact of the results of a high-stakes assessment is likely to be more powerful than that arising from more informal or lower-stakes assessments.

Although little is known about feedback after high-stakes assessments, it is relevant to consider learning-related characteristics, such as goal orientation, control of learning, self-efficacy and motivation, which influence feedback in other settings.

The role of goal orientation is important in understanding students' use of feedback. Some individuals have a mastery or learning goal orientation, to develop competence and mastery of new skills, while others have performance goal orientation, to demonstrate their competence to others, seek positive comments and avoid negative comments about their work.¹³ A student with a learning goal orientation is likely to regard feedback as a diagnostic tool providing useful information to help him acquire greater competence, whereas a student with a performance goal orientation is likely to view feedback as a judgement about himself as well as his competence.¹⁴ Negative feedback can be particularly demoralising as it is in direct opposition with the goal of appearing competent. Performance goal orientation can be subdivided into two main categories: approach and avoid goals.¹⁵ Performance-approach oriented students are concerned with demonstrating that they are more competent than others, whereas performance-avoid oriented students are keen to avoid appearing incompetent. In the workplace, medical trainees with a higher learning goal orientation perceive many benefits and few costs from feedback, whereas those with a higher performance goal orientation perceive more costs from feedback.¹⁶ We therefore hypothesised that students with a learning goal orientation would make greater use of feedback information after summative assessment compared with students with a performance orientation. Further, we hypothesised that students with a performance approach goal orientation who performed badly would make less use of feedback than those students who performed well, whereas performance-avoid oriented students would make more use of feedback if they performed badly, as they would be concerned about potential failure.

Motivation influences the effort taken in order to achieve a goal.¹⁷ Intrinsic motivation refers to a desire within an individual to succeed, rather than being reliant on external factors, whereas extrinsic motivation is focussed on the achievement of outcomes in order to satisfy external pressures. Students with higher intrinsic motivation, focussed towards achievement, invested greater time in studying.¹⁸ Intrinsic motivation correlates positively to time spent reflecting on learning.¹⁹ Therefore we hypothesised that students with higher intrinsic motivation would view more feedback information than students with higher extrinsic motivation.

‘Control of learning’ - students’ beliefs that their academic achievements are contingent on their own efforts, not on external factors - is another factor to consider.²⁰ Students who believe that their own efforts will affect their achievement put more effort into their learning.²⁰ We therefore expected that students with a stronger belief in the control of their learning would be more likely to make use of feedback.

Successful academic achievement can lead to an improvement in self-efficacy, which is concerned with a person’s belief in their ability to succeed in a particular situation.²¹ A higher self-efficacy, once developed, can help students to cope with negative feedback.²² We therefore predicted that medical students with higher self-efficacy would make greater use of feedback.

The term ‘self-presentation of low achievement’, refers to students’ preferences for keeping their achievements hidden from their peers.²³ Students who under-achieve compared with their peers score higher on this concept.²⁴ An unwillingness to be open about performance may impair students’ willingness to engage with feedback as they may not wish to acknowledge to others areas where they need to improve, though empirical data on this effect are lacking. We therefore hypothesised that students who preferred to keep their achievements hidden would make less use of the feedback.

Feedback can be valuable to learners as it provides information to help them meet their goals and it can help to increase their self-confidence.¹⁰ On the other hand feedback can present risks to learners if it provides information which does not fit in with their own self-image. Feedback-seeking behaviour is mediated by the learners’ perception of the likely value and risk of feedback.¹⁶ In particular, learners who regard feedback as valuable are, unsurprisingly, more likely to seek feedback. Because perception of feedback value tends to correlate inversely to perception of feedback risk, learners who regard feedback as more risky are likely to seek less feedback, although empirical results are equivocal.¹⁶ Therefore we hypothesised that students who perceive a higher value of feedback will make more use of feedback information after summative assessment, whereas those who regard it as more risky will make less use of the feedback provided.

Learners are not simply passive recipients of feedback, but seek out the feedback using both direct and indirect strategies.⁸ In the former case, learners ask feedback-providers directly, who in turn share only the feedback that they wish to share. By contrast, employing an indirect strategy means that the feedback-seeker attempts to infer the feedback by monitoring reactions of others (for instance, the non-verbal behaviour of a supervisor). When designing feedback systems for medical students, it would be useful to understand to what extent direct and indirect feedback-seeking strategies are employed. We hypothesised that more active seekers of feedback in the workplace would also be more active seekers of feedback after summative assessment, and a similar relationship would be expected from those learners who were feedback monitors.

Once feedback has been obtained, it can sometimes raise as many questions as it answers, especially if the feedback is negative or unexpected. This is defined as feedback uncertainty.²⁵ Learners who perceive higher levels of feedback uncertainty tend to be less active seekers of feedback.²⁵ We therefore hypothesised that students who rated more highly on feedback uncertainty would make less use of the feedback information provided.

Although the literature suggests that comments from examiners are appreciated by students, it is often not practical to provide them in the context of a busy, high-stakes OSCE with very limited time between candidates. As a result, summative assessments often involves the awarding of grades or numerical marks, but the use of grades when giving feedback has been criticised.²⁶ This stems from studies demonstrating that awarding grades blocks learners' receptivity to formative comments.²⁷ However, this criticism primarily derives from studies that have been conducted on schoolchildren or in the context of formative, low-stakes assessments. The effect of giving grades, or benchmarking a student's mark against others in the same cohort, has not previously been explored in the context of summative assessment in medicine. Benchmarking can help students to self-assess their own performance with a greater degree of accuracy.²⁸ One study of medical students offered feedback after the awarding of grades for a summative written assignment has demonstrated that higher performers were more likely to access the feedback than students who had performed less well or failed.⁷ Performance in assessment may therefore influence receptivity to feedback. We are not aware of any studies comparing students' summative OSCE performances (and the awarding of marks) with their subsequent engagement with feedback. There is therefore a need to understand the influence of performance on the use of feedback after summative assessment.

This study investigates how students receive and use feedback in the context of summative assessment. The aim was to explore how students receive feedback (in the form of profile scores on OSCE stations and benchmarking, delivered via a website) after a summative assessment. Specifically, we intended to answer the following questions:

1. To what extent do medical students engage with feedback after a high-stakes OSCE?
2. Do goal orientation, motivation, self-efficacy, control of learning beliefs or students' attitudes to the benefits or risks of feedback influence how students engage with the feedback?
3. Does performance in the current or previous assessment affect engagement with feedback?

Methods

Context

The study took place at Keele University School of Medicine in 2011. The third year (out of five), comprising 138 students, represents a transition to much more clinically-based learning with 80% of learning taking place in the clinical environment. At the end of the year, students have a summative 12-station OSCE, in which they have to pass at least 8 stations. Students have to pass to progress to Year 4. Each station is 10 minutes long and is marked with generic consultation rating scales that are used across multiple stations.²⁹ Students who failed the OSCE were required to discuss their performance with a senior tutor, although they were not required to view the website either before or after this meeting. This took place at least 6 weeks after the results were released. Students who passed the OSCE were not offered this opportunity.

Feedback Website

A website was developed within the virtual learning environment to deliver feedback about the OSCE in various formats: station-by-station or on skills across stations. In the station-by-station manner students could view information in various ways: ‘pass-fail’ webpages, which gave information about how close to the pass mark and how far short of the maximum mark students were; ‘global’ webpages presented the examiners’ global mark with a graphical comparison to the cohort average for that station; ‘skills breakdown’ pages presented a breakdown of the different skills assessed at that station, with a graphical comparison to the cohort average; ‘detailed comparison’ webpages presented the cohort’s performance on that station in a frequency-distribution chart, with additional information about the pass mark and the student’s individual performance. Students could also look across stations at their performance on skills which were assessed at multiple stations. Skills that were assessed in more than one station were filtered out and presented in numerical form (the skills summary) or graphically compared with the cohort (cross-station breakdown). Students could also access ‘Next steps’ webpages which provided more guidance on reflecting and responding to the feedback (with specific worksheets to download and work through), along with encouragement to develop action plans based on the well-known SMART format (Specific, Measurable, Achievable, Realistic, Timely).³⁰ A table showing a summary of the 130 different webpages available is shown in Appendix S1. Results from the OSCE were uploaded to the website via an Excel spreadsheet. Four hours after results of the OSCE were delivered, the feedback was released. Use of the website was left entirely to students’ discretion; there was no requirement to log on. The website recorded the total pages viewed, time spent on the website, and page hits for each area. Data were collected on students’ OSCE performance, performance in the previous year’s 20-station practical skills assessment objective structured skills examination (OSSE) and on students’ performance in the written assessments for that year.

Questionnaire

A 51-item questionnaire was developed to measure learning-related characteristics and attitudes to feedback. It was based on scales of previously validated instruments that were relevant to feedback in a summative context and had an acceptable reliability. Some items were slightly reformulated to better fit the context of the study. Some scales were shortened. Full details are shown in Appendix S2. Scales selected covered goal orientation (mastery, performance approach and performance avoid), intrinsic and extrinsic motivation, control of learning beliefs, self-efficacy for learning and performance, attitudes towards self-presentation of low achievement, attitudes towards the value, uncertainty or risk of feedback and self-reported frequency of direct or indirect feedback-seeking behaviour.^{20,23,25, 31-34} Responses were recorded on a 5-point Likert scale.

Data Collection

Questionnaires were given out at the end of a lecture two weeks before the summative OSCE. Students were told that this was part of a research study and that they were free to consent or not. Participants were given a chocolate bar as a token of thanks. Students' library card numbers were recorded to allow linking of responses to examination performances and use of the website. Usage of the website was monitored for 8 weeks following the release of the feedback in order to record repeat visits to the website and to record if some students first visited the website after a significant delay.

Data Analysis

To answer the first research question, website usage was analysed by calculating the number of separate webpage 'hits' for each student and by calculating the number of separate visits to the website made by each student. The total time (from login to logout) spent on each visit was also recorded. Descriptive statistics were obtained using SPSS Version 18 [SPSS Inc, Chicago, IL, USA]. Latent class analyses (LCA; WINMIRA software³⁵) were used to investigate different profiles of students using the feedback information. LCA is a statistical probabilistic method that clusters subgroups of students according to similar profiles.³⁶ It has previously been used successfully to cluster homogenous groups of students together when studying learning environments and enables an evaluation of the interaction between student profiles and other learning characteristics.³⁷ This method uses the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Consistent Akaike Information Criterion (CAIC) as indices for the quality of the computed solutions. Lowest values on these criteria indicate best fit to the data. Empirical distributions were generated using the bootstrapping method (Cressie-Read and Pearson). Based on these indices the best fitting solution was selected, representing the optimal number of profiles to be used to describe the data. The variables entered in the analyses for defining the profiles were the total webpages visited, the number of separate visits to the website, the average number of webpages viewed per minute, and the number of webpages viewed for each section of the website (pass-fail, global, skills breakdown, detailed comparisons, next steps, skills summary and cross-station breakdown.

To answer the second research question, descriptive statistics were calculated for each of the learning-related characteristic scales. One-way analyses of variance (ANOVAs) were performed to explore the relationship between learning-related characteristics and use of the feedback (as defined by the number of separate visits students made to the website). Post-hoc analyses were performed with the Tukey's Honestly Significant Difference Test. To answer the third research question, ANOVAs and χ^2 were performed to explore the relationship between OSCE performance (in the current year) and use of the feedback, and between OSSE performance (in the previous year) and use of the feedback.

Ethical Approval

Ethical approval was obtained via the Keele School of Medicine Ethics Committee before the study commenced.

Results

Use of the Website

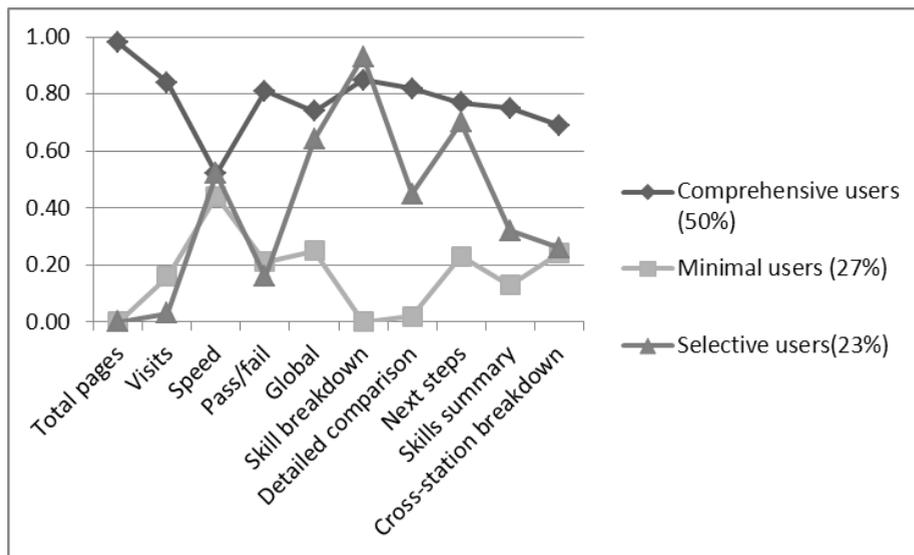
The website was viewed by 132 students (95.7%). Of those who viewed it, 115 (87.1%) first visited the website on the first day that the results became available; 5 (4.3%) waited more than 2 weeks before first viewing the website. The number of separate visits ranged from 1 to 5 (mean 1.90, SD 1.17, median 1); 65 students (49.24%) visited the website more than once. The number of pages viewed ranged from 2 to 377 (mean 123.47, SD 74.24, median 111). There were 130 different webpages available. Descriptive statistics for use of the different aspects of the website are shown in Table 1. Pass-fail webpages were visited very frequently, while other breakdowns of marks were only visited about half as frequently. Many students viewed the webpages quickly, with half viewing at least 6 webpages per minute.

To select the best-fitting latent class analysis model, the solutions for 1 - 10 classes were computed (Table S1, see supplementary information). On the basis of the BIC and CAIC, the 2-class solution is favoured as these indices are at their lowest, suggesting the best fit. However, based on the AIC and the bootstraps the 4-class model is preferred. The 3-class solution provided compromise and the best fit overall to the data. The three classes of students differed in how they utilised the website. For ease of reference, these are described as 'Comprehensive users', 'Selective users' and 'Minimal users' (see Figure 1). The Comprehensive users, representing 50% of students, made high use of the website across all areas. The Minimal users (27% of students) made low use of the website across all areas. The Selective users (23% of students) made less use of the website overall. Compared with the Comprehensive users, they made fewer visits and looked at fewer pass-fail, skills summary and cross-station breakdown pages, but made similar use of the global, skills breakdown and next steps sections.

Table 1: Use of different parts of website

	Mean	Median	Standard Deviation	Lower Quartile	Upper Quartile	Interquartile Range
Total webpages visited	123.47	111	74.24	70	170	99
Total number of visits to website	1.90	1	1.18	1	2	1
Speed (average pages viewed per minute)	6.56	6	2.68	4.69	8.12	3.43
Pass/fail webpages viewed	25.14	20	18.61	13	32	19
Global mark webpages viewed	15.23	13	12.12	6	19	13
Skills breakdown webpages viewed	15.61	13	16.79	9	17	8
Detailed comparison with other students webpages viewed	13.86	13	13.37	3	18	15
Skills summary webpages viewed	17.86	12	17.90	3	28	25
Station breakdown webpages viewed	10.01	4	13.99	2	12	10
Next steps webpages viewed	1.43	1	1.75	0	2	2

Figure 1: The 3-class model for latent class analysis



Learning-related Characteristics and Use of Feedback

The questionnaire was completed by 113 students (81.9%); 69 (61%) were female which is comparable to their proportion in the year as a whole (59%). Descriptive statistics for the questionnaire scales are shown in Table 2, along with the Cronbach’s alpha per scale. The scales showed moderate to acceptable internal consistencies (alphas ranging from .52 to

.80). Students on average scored above the mid-point of the scale on mastery and performance avoid goal orientation, intrinsic motivation, control of learning beliefs, self-efficacy, value of feedback and frequency of feedback monitoring; they scored below the mid-point for performance approach goal orientation, self-presentation of low achievement, risk in feedback seeking and feedback-related uncertainty (one-sample *t* tests, $p < 0.01$; see Table 3). ANOVAs demonstrated that students who were Comprehensive users scored more highly on the Value of Feedback scale than Selective users ($F(2, 110) = 3.42$; $p < 0.05$; $\Delta = .34$, $p < 0.05$, $d = 0.57$). Minimal users scored more highly than Selective users on the Extrinsic Motivation scale ($F(2, 110) = 3.706$; $p < 0.05$; $\Delta = 0.58$, $p < 0.05$, $d = 0.67$). Post-hoc analysis of the ANOVA comparing learning-relating characteristics with the number of separate website visits did not demonstrate significant differences.

Performance and Use of Feedback

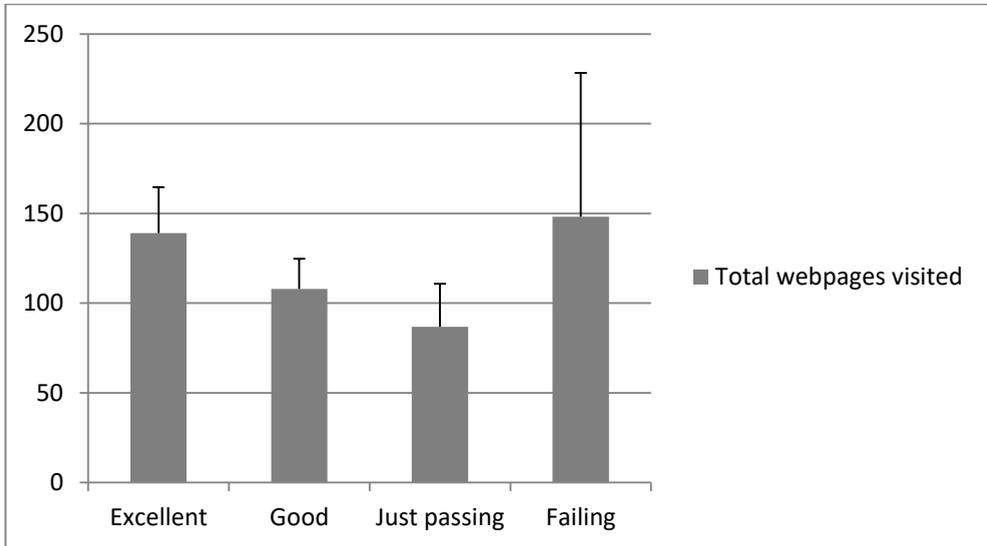
An ANOVA demonstrated that the Comprehensive users had a significantly higher OSCE ($F(2, 134) = 4.65$; $p < 0.05$; $\Delta = 13.02$, $p < 0.05$, $d = 0.57$) and OSSE score ($F(2, 134) = 5.21$; $p < 0.01$; $\Delta = 5.04$, $p < 0.01$, $d = 0.58$) than the Minimal users. Inspecting the means plots of stations passed against total website usage indicated that there was no direct linear relationship between overall OSCE score and website usage but a U-shaped relationship existed. To facilitate further analysis, students were grouped into categories: Excellent (passed all 12 stations, $N = 47$), Good (failed 1-2 stations, $N = 64$), Just-passing (failed 3-4 stations, $N = 17$) and Failing (failed 5 or more stations, $N = 9$). The results are shown graphically in Figures 2 and 3. Within the different categories of OSCE performance, there was a trend towards Excellent students viewing the highest number of webpages, and Just-passing students viewing the website the least ($F(3,133) = 3.05$; $p < 0.05$; $\Delta = 52.18$, $p < 0.10$, $d = 0.75$). As can be seen in Figure 4, fourteen (29.79%) of the Excellent students visited the website 3 or more times; no Just-passing students visited the website more than twice (likelihood ratio = 29.85, $df = 15$, $p < 0.01$). Excellent students viewed more global webpages than Just-passing students ($F(3,133) = 4.67$; $p < 0.01$; $\Delta = 10.27$, $p < 0.05$, $d = 0.97$) and Failing students ($\Delta = 11.68$, $p < 0.05$, $d = 1.18$). Excellent students viewed more detailed comparison webpages than all other students ($F(3,133) = 6.20$; $p < 0.01$; Good: $\Delta = 7.74$, $p < 0.05$, $d = 0.55$; Just-passing: $\Delta = 11.77$, $p < 0.01$, $d = 0.89$; Failing: $\Delta = 13.85$, $p < 0.05$, $d = 1.09$). Failing students viewed more skill breakdown pages than Good students ($F(3,133) = 3.41$; $p < 0.05$; $\Delta = 17.84$, $p < 0.05$, $d = 2.63$) or Just-passing students ($F(3,133) = 3.41$; $p < 0.05$; $\Delta = 19.18$, $p < 0.05$, $d = 2.81$). Failing students used the next steps webpages more frequently than Good students ($F(3,133) = 5.81$; $p < 0.01$; $\Delta = 2.32$, $p < 0.01$, $d = 0.84$) or Just-passing students ($\Delta = 2.56$, $p < 0.01$, $d = 0.96$). There was no significant relationship between performance in the written assessments and use of the website. No significant differences were found in the use of other areas of the website.

Table 2: Descriptive statistics for learning-related student characteristics in questionnaire

Learning-related student characteristics	Items in scale	Mean	Standard deviation	One-sample t test*	Cronbach's alpha per scale	Sample item (all items shown in Appendix S2)
Mastery goal orientation	5	4.47	0.45	p<0.01	0.64	One of my goals this year is to learn as much as I can
Performance approach goal orientation	5	2.70	0.78	p<0.01	0.80	It's important to me that other students in my year think I am good at my work
Performance avoid goal orientation	3	3.50	0.74	p<0.01	0.68	One of my goals is to avoid looking like I have trouble doing the work
Intrinsic motivation towards accomplishment	4	3.38	0.72	p<0.01	0.71	I go to medical school because it allows me to experience a personal satisfaction in my quest for excellence in my studies
Extrinsic motivation introjected regulation	4	2.99	0.82		0.69	I go to medical school because of the fact that when I succeed in medical school I feel important
Control of learning beliefs	4	3.77	0.66	p<0.01	0.64	If I study in appropriate ways, then I will be able to learn the material in this course
Self-efficacy for learning and performance	6	3.37	0.60	p<0.01	0.75	I'm confident I can do an excellent job on the assignments and tests in this course
Self-Presentation of Low Achievement	4	2.45	0.68	p<0.01	0.52	If other students found out I did well on a test, I would tell them it was just luck even if that wasn't the case
Value of feedback	3	4.31	0.56	p<0.01	0.52	It is important to me to receive feedback on my performance
Risk in feedback seeking	5	2.17	0.69	p<0.01	0.78	It is not a good idea to ask your tutor for feedback; he/she might think you are incompetent
Feedback related uncertainty	4	2.67	0.79	p<0.01	0.79	When my tutors give me feedback about my performance, I am not immediately sure what it requires me to do
Frequency of feedback monitoring	2	3.49	0.83	p<0.01	0.76	How frequently do you observe what performance behaviours and skills your tutor rewards and use this as feedback on your own performance?
Frequency of active feedback seeking	1	3.18	1.09			How frequently do you seek feedback from your tutors about your performance?

*One-sample t-test compares the mean for the scale with the mid-point of the scale (3)

Figure 2: Mean number of webpages by objective structured examination performance (OSCE) performance



Note: error bars represent 95% confidence intervals

Figure 3: Mean number of webpages visited by section and objective structured examination (OSCE) performance

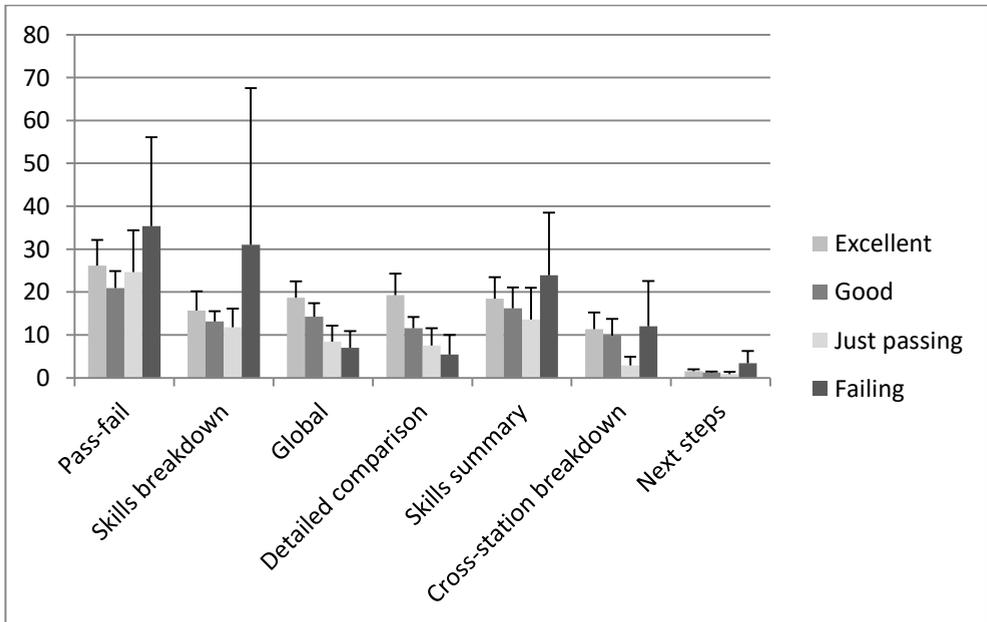
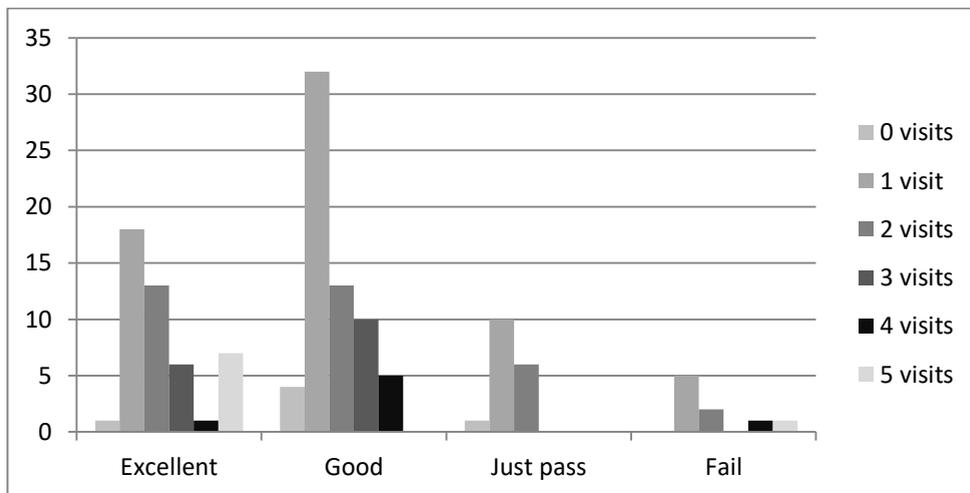


Figure 4: Number of visits to website by objective structured examination (OSCE) performance

Discussion

This study demonstrated that the majority of students will engage with feedback after a summative OSCE, delivered via a website. Latent class analysis demonstrated that there were three broad groups of students who used the website in different ways. About half the students seemed very interested in the website information, a quarter showed relatively little interest and a quarter used the website in a more selective way. Unsurprisingly, the Comprehensive users scored more highly on the Value of Feedback scale. The Minimal users scored more highly for Extrinsic Motivation. Students who passed all OSCE stations made most use of the website, whereas students who just passed made least use. No student in this latter category visited the website more than twice.

Unlike other studies which simply demonstrated whether or not the feedback was collected, this study provides the opportunity to understand how often the students return to the feedback and to see which areas of feedback different subgroups of students preferentially view. Interestingly, the strongest performers tended to make more use of aspects of the website that compared their performance with other students (global webpages, detailed comparison with other students). Although these sections show them how they are doing compared with their peers, they do not demonstrate why they have performed strongly; these aspects are better shown in the sections focusing on individual skills, which are not viewed more often by these students. This suggests that these students are perhaps looking for positive affirmation or reassurance from the feedback, rather than on how to use the feedback in an educational way for the future. This fits with previous qualitative work on feedback, which found implicit evidence that students and trainees were seeking feedback to build their confidence rather than to change their behaviour.¹⁰ The small number of

students who failed the OSCE made considerable use of the website, although there was substantial variability between students. These students are all invited to discuss their results with a senior tutor, so it is possible this could have influenced their use of the website.

Intriguingly, the students who just passed the OSCE made least use of the feedback, yet they are at risk of failing future assessments and arguably have most to gain from the feedback. This group is unlikely to be homogeneous, but will comprise students who often just pass, others who normally do much better and others who failed previous assessments. It is therefore surprising that these students behaved in a relatively homogeneous way by visiting the website no more than twice, unlike other types of students who showed a more varied range of engagement with the feedback. It is possible that the comparative data with other students was not so reassuring for this group and therefore there was less incentive for them to review this information. There is a clear need to study this subset of students in more detail to see if it is possible to design interventions which would increase their engagement with feedback. Although there has been considerable focus on students who fail OSCEs, the group of students who just pass the assessments have been largely neglected. Given that there are potentially many more students in this category, this lack of scrutiny cannot be justified.

It was surprising that no clear relationship was found between the students' learning-related characteristics and their use of feedback as relationships have previously been demonstrated in the literature. It is unclear whether this is because the feedback was delivered in the context of a summative rather than formative assessment, or because other factors are more important. Others have suggested that the role of cognitive learning styles in affecting performance may not be as straightforward as previously thought.³⁸ This requires further study.

There are several limitations to our study. It was done within a single year group in a single medical school, so it is not certain to what extent the results will be generalisable to other medical schools, especially those with a different assessment regime. The feedback provided was a numerical breakdown of scores and skills but did not include examiners' comments for reasons of feasibility. It is possible that a different pattern may have emerged if comments were also included, as there is evidence that students may engage better with this type of feedback.³⁹ As this was a quantitative study it cannot differentiate students who intensely engaged with the feedback from those who simply looked at the screen without attempting to understand their performance, although the frequency with which some students returned to the website provides circumstantial evidence. It also cannot explain why students used the website in the way that they did. A qualitative study would help with this.

The time when the feedback is administered may be relevant. It is recognised that feedback should be delivered to students at a time when it is needed, and when they have an opportunity to make use of it.² As these students will not have another OSCE for 12 months, it could be argued that they do not have an immediate need to make use of the feedback. However, although this was a summative OSCE, it did not occur at the end of the academic year, but was two-thirds of the way through the year, and the students did then have some clinical attachments, which provided an opportunity for them to make use of the feedback and adapt their clinical skills.

Our study supports the provision of feedback after summative assessments as the majority of students will engage with the feedback. If we are to meet our goal of encouraging students towards optimal competence, then it is important to understand the group of students who just pass assessments, as they engage less well with feedback. Once we understand this group better, then interventions can be planned to improve their engagement with feedback and, hopefully, their subsequent performance.

References

1. Dochy F, Segers M, Gijbels D, Struyven, K. Assessment Engineering: Breaking down barriers between teaching and learning, and assessment. In: Boud D, Falchikov N, (Eds.), *Rethinking assessment in higher education: learning for the longer term*. Oxford: Routledge; 2007. p 87-100.
2. Shute V. Focus on formative feedback. *Review of Educational Research* 2008; 78: 153-89.
3. Brookhart SM. Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice* 2001; 8: 153-169.
4. Wass V, Van der Vleuten C, Shatzer J, Jones R. The assessment of clinical competence. *The Lancet* 2001; 357: 945-949.
5. Norcini JJ. Setting standards on educational tests. *Medical Education* 2003; 37: 464-469.
6. Kluger AN, DeNisi A. The effects of feedback interventions on performance; a historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin* 1996; 119: 254-284.
7. Sinclair HK, Cleland, JA. Undergraduate medical students: who seeks formative feedback? *Medical Education* 2007; 41: 580-582.
8. Ashford SJ, Blatt R, VandeWalle D. Reflections on the looking glass: a review of research on feedback-seeking behaviour in organizations. *Journal of Management* 2003; 29: 773-799.
9. Mann K, Van der Vleuten C, Eva K, Armson H, Chesluk B, Dornan T, Holmboe E, Lockyer J, Loney E, Sargeant J. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Academic Medicine* 2011; 86: 1120-1127.
10. Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012; 17: 15-26.
11. Nicol, D. From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education* 2010; 35: 501-517.

12. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education* 2008; 13:275–288.
13. Dweck CS, Leggett EL. A social-cognitive approach to motivation and personality. *Psychological Review* 1988; 95: 256-273.
14. VandeWalle D. A goal orientation model of feedback-seeking behaviour. *Human Resource Management Review* 2003; 13: 581-604.
15. VandeWalle D. Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement* 1997; 8: 995-1015.
16. Teunissen PW, Stapel DA, Van der Vleuten CPM, Scherpbier AJJA, Boor K, Scheele F. Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine* 2009; 84: 910-917.
17. Maslow AH. A theory of achievement motivation. In Harriman PL (Ed). *Twentieth century psychology: recent developments*. Manchester: Ayer Publishing 1970; 22-48.
18. Wilkinson TJ, Wells JE, Bushnell JA. Medical student characteristics associated with time in study: is spending more time always a good thing? *Medical Teacher* 2007; 29:106-110.
19. Sobral DT. What kind of motivation drives medical students learning quests? *Medical Education* 2004;38: 950-957.
20. Zusho A, Pintrich PR, Coppola B. Skill and will: the role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education* 2003;25:1081-1094.
21. Bandura A. Self-efficacy: toward a unifying theory of behavioural change. *Psychological Review* 1977; 84:191-215.
22. Bandura A, Locke EA. Negative Self-Efficacy and Goal Effects Revisited. *Journal of Applied Psychology* 2003; 88: 87–99.
23. Midgley C. (2000). *Manual for the Patterns of Adaptive Learning Scales*. University of Michigan. Available online at http://www.umich.edu/~pals/PALS%202000_V12Word97.pdf [Accessed 24 March 2013.]
24. Ketter LC. High-stakes testing, achievement-goal structures, academic related perceptions, beliefs, strategies, and school belonging among selected eighth-grade students in a northwest Florida school district. Education doctorate University of West Florida 2006. http://etd.fcla.edu/WF/WFE0000029/Ketter_Lynn_Carol_200605_EdD.pdf [Accessed 24 March 2013.]
25. Fedor DB, Rensvold, RB, Adams, SM. An investigation of factors expected to affect feedback seeking: A longitudinal field study. *Personnel Psychology* 1992; 45: 779-805.
26. Black P, Wiliam D. *Assessment and Classroom Learning*. *Assessment in Education: Principles Policy & Practice* 1998; 5: 7-74.
27. Butler R. Task-involving and ego-involving properties of evaluation: effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology* 1987; 79: 474-478.
28. Srinivasan M, Hauer K, Der-Martirosian C, Wilkes M, Gesundheit N. Does feedback matter? Practice-based learning for medical students after a multi-institutional clinical performance examination. *Medical Education* 2007; 41: 857–865.
29. Lefroy J, Gay S, Gibson S, Williams S, McKinley RK. Development and face validation of an instrument to assess and improve clinical consultation skills. *International Journal of Clinical Skills* 2011; 5: 115-125.

30. Doran GT. There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review* 1981; 70, 35-36.
31. Anderman EM, Urdan T, Roeser R. (2003). The Patterns of Adaptive Learning Survey: history, development and psychometric properties. Paper prepared for Indicators of Positive Development Conference, 2003. available online at http://www.childtrends.org/Files/Child_Trends-2003_03_12_PD_PDConfAUR.pdf. [Accessed 24 March 2013.]
32. Wolters CA, Yu SL, Pintrich PR. The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences* 1996; 8: 211-238.
33. Ashford SJ. Feedback-seeking in individual adaptation: a resource perspective. *Academy of Management Journal* 1986; 29: 465-487.
34. Vallerand RJ, Pelletier LG, Blais MR, Briere NM, Senecal C, Vallieres EF. The Academic Motivation Scale: a measure of intrinsic, extrinsic and amotivation in education. *Educational and Psychological Measurement* 1992; 52: 1003-1017.
35. Davier vM. WINMIRA - A program system for analyses with the Rasch-model, with the latent class analysis and with the mixed-Rasch model. Kiel: IPN; 1999.
36. Lazarsfeld PF, Henry NW. *Latent structure analysis*. Boston: Houghton Mifflin Co. 1968.
37. Seidel T. The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research* 2006; 9: 253-271.
38. Cook DA. Revisiting cognitive and learning styles in computer-assisted instruction: not so useful after all. *Academic Medicine* 2012; 87: 778-784.
39. Lipnevich AA, Smith JK. Effects of Differential Feedback on Students' Examination Performance. *Journal of Experimental Psychology: Applied* 2009; 15: 319-333.

Supplementary Information

Table S1: Indices for 1 – 10 LCA Solutions, With Ten Measures With Respect to Website Usage

Class	Log L	Parameters	Information Criteria			Bootstraps	
			AIC	BIC	CAIC	Cressie-Read	Pearson
		(k)					
1	-944.61	10	1909.22	1938.49	1948.49	0	0
2	-786.04	21	1614.08	1675.55	1696.55	0.025	0.1
3	-763.37	32	1590.74	1684.41	1716.41	0.055	0.123
4	-744.52	43	1575.05	1700.92	1743.92	0.27	0.463
5	-735.37	54	1578.75	1736.82	1790.82	0.113	0.258
6	-723.98	65	1577.96	1768.23	1833.23	0.178	0.305
7	-720.32	76	1592.65	1815.12	1891.12	0.123	0.213
8	-713.39	87	1600.78	1855.45	1942.45	0.128	0.225
9	-709.46	98	1614.91	1901.78	1999.78	0.098	0.198
10	-699.22	109	1616.43	1935.5	2044.5	0.145	0.213

Appendix S1: Table showing breakdown of webpages

Type of webpage	Number of pages
Main menu	1
Pass-fail webpages for questions 1-12	12
Skills breakdown for questions 1-12	12
Global mark for questions 1-12	12
Detailed comparison for questions 1-12	12
Skills summary for each of 39 skills	39
Station breakdown for each of 39 skills	39
Next steps including options to download handouts	3
Total	130

Appendix S2: Table showing the changes to the questionnaire items

<i>Original question</i>	<i>Revised question</i>
Mastery Goal Orientation – Patterns of Adaptive Learning Scales (PALS)²³	
<i>Alpha: 0.85</i>	<i>Alpha: 0.64</i>
It's important to me that I learn a lot of new concepts this year.	It's important to me that I learn a lot of new concepts this year.
One of my goals in class is to learn as much as I can.	One of my goals this year is to learn as much as I can.
One of my goals is to master a lot of new skills this year.	One of my goals is to master a lot of new skills this year.
It's important to me that I thoroughly understand my classwork	It's important to me that I thoroughly understand my work
It's important to me that I improve my skills this year.	It's important to me that I improve my skills this year.
Performance-Approach Goal Orientation - PALS²³	
<i>Alpha: 0.89</i>	<i>Alpha: 0.80</i>
It's important to me that other students in my class think I am good at my classwork.	It's important to me that other students in my year think I am good at my work.
One of my goals is to show others that I'm good at my classwork.	One of my goals is to show others that I'm good at my work.
One of my goals is to show others that classwork is easy for me.	One of my goals is to show others that work is easy for me.
One of my goals is to look smart in comparison to the other students in my class.	One of my goals is to look smart in comparison to the other students in my year.
It's important to me that I look smart compared to others in my class.	It's important to me that I look smart compared to others in my year.
Performance-Avoid Goal Orientation - PALS²³	
<i>Alpha: 0.74</i>	<i>Alpha: 0.68</i>
It's important to me that I don't look stupid in class.	It's important to me that I don't look stupid in comparison with others.
One of my goals is to keep others from thinking I'm not smart in class.	omitted
It's important to me that my teacher doesn't think that I know less than others in class.	It's important to me that my teacher doesn't think that I know less than others in my group.
One of my goals in class is to avoid looking like I have trouble doing the work.	One of my goals is to avoid looking like I have trouble doing the work.
Intrinsic motivation towards accomplishment - Academic Motivation Scale (AMS)³⁴	
<i>Alpha: 0.85</i>	<i>Alpha: 0.71</i>
Why do you go to college?	
For the pleasure I experience while surpassing myself in my studies	I go to medical school for the pleasure I experience while excelling in my studies
For the pleasure that I experience while I am surpassing myself in one of my personal accomplishments	I go to medical school for the pleasure I experience while excelling in one of my personal achievements
For the satisfaction I feel when I am in the process of accomplishing difficult academic activities	I go to medical school for the satisfaction I feel when I am in the process of achieving difficult academic activities
Because college allows me to experience a personal satisfaction in my quest for excellence in my studies	I go to medical school because it allows me to experience a personal satisfaction in my quest for excellence in my studies

Extrinsic motivation – introjected – AMS³⁴	
<i>Alpha: 0.84</i>	<i>Alpha: 0.69</i>
Why do you go to college?	
To prove to myself that I am capable of completing my college degree	I go to medical school to prove to myself that I am capable of completing my degree
Because of the fact that when I succeed in college I feel important	I go to medical school because of the fact that when I succeed in medical school I feel important
To show myself that I am an intelligent person.	I go to medical school to show myself that I am an intelligent person
Because I want to show myself that I can succeed in my studies	I go to medical school because I want to show myself that I can succeed in my studies
Control of Learning Beliefs - MSLQ	
<i>Alpha 0.68</i>	<i>Alpha: 0.64</i>
If I study in appropriate ways, then I will be able to learn the material in this course.	If I study in appropriate ways, then I will be able to learn the material in this course.
It is my own fault if I don't learn the material in this course.	It is my own fault if I don't learn the material in this course.
If I try hard enough, then I will understand the course material.	If I try hard enough, then I will understand the course material.
If I don't understand the course material, it is because I didn't try hard enough	If I don't understand the course material, it is because I didn't try hard enough
Self-efficacy for Learning and Performance - MSLQ	
<i>Alpha 0.93</i>	<i>Alpha: 0.75</i>
I believe I will receive an excellent grade in this class.	I believe I will receive excellent marks this year.
I'm certain I can understand the most difficult material presented in the readings for this course.	I'm certain I can understand the most difficult material presented in this course.
I'm confident I can learn the basic concepts taught in this course.	I'm confident I can learn the basic concepts taught in this course.
I'm confident I can understand the most complex material presented by the instructor in this course.	<i>omitted</i>
I'm confident I can do an excellent job on the assignments and tests in this course.	I'm confident I can do an excellent job on the assignments and tests in this course.
I expect to do well in this class.	<i>omitted</i>
I'm certain I can master the skills being taught in this class.	I'm certain I can master the skills being taught in this year.
Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.	Considering the difficulty of this course, the teachers, and my skills, I think I will do well in this year.
Self-Presentation of Low Achievement – PALS²³	
<i>Alpha: 0.78</i>	<i>Alpha: 0.52</i>
I would avoid participating in a class if it meant that other students would think I know a lot.	I would avoid participating in a teaching session if it meant that other students would think I know a lot.
If other students found out I did well on a test, I would tell them it was just luck even if that wasn't the case.	If other students found out I did well on a test, I would tell them it was just luck even if that wasn't the case.
I wouldn't volunteer to answer a question in a class if I thought other students would think I was smart.	<i>omitted</i>
If I did well on a school assignment, I wouldn't want other students to see my grade.	If I did well on an assignment, I wouldn't want other students to see my marks.
It's very important to me that I don't look smarter than others in my class.	It's very important to me that I don't look smarter than others in my year.

Chapter 2

If I were good at my class work, I would try to do my work in a way that didn't show it.	<i>omitted</i>
One of my goals in class is to avoid looking smarter than other kids.	<i>omitted</i>
Risk in feedback seeking ³³	
<i>Alpha 0.80</i>	<i>Alpha: 0.78</i>
I think my boss would think worse of me if I asked him/her for feedback.	I think my tutor would think worse of me if I asked him/her for feedback.
I would not be nervous about asking my boss how he/she evaluates my behaviours and skills. (R)	I would not be nervous about asking my tutor how he/she assesses my behaviours and skills. (R)
It is not a good idea to ask your co-workers for feedback; they might think you are incompetent.	<i>omitted</i>
It is embarrassing to ask my co-workers for their impression of how I am doing at work.	<i>omitted</i>
It would not bother me at all to ask my boss for feedback. (R)	It would not bother me at all to ask my tutor for feedback. (R)
It is not a good idea to ask your boss for feedback; he/she might think you are incompetent.	It is not a good idea to ask your tutor for feedback; he/she might think you are incompetent.
It is embarrassing to ask my tutor boss for feedback.	It is embarrassing to ask my tutor for feedback.
I think my co-workers would think worse of me if I asked them for feedback.	<i>omitted</i>
I would not be nervous about asking my co-workers how they evaluate my behaviors. (R)	<i>omitted</i>
It is better to try and figure out how you are doing on your own rather than ask your co-workers for feedback.	<i>omitted</i>
Value of Feedback ³³	
<i>Alpha 0.72</i>	<i>Alpha: 0.52</i>
It is important to me to receive feedback on my performance.	It is important to me to receive feedback on my performance.
I would like to get more feedback on what behaviours will help me do better in performing my job.	I would like to get more feedback on what behaviours and skills will help me do better as a medical student.
I find feedback on my performance useful.	I find feedback on my performance useful.
It is important for me to receive feedback on my potential for advancement within (X).	<i>omitted</i>
I would like to get more feedback on what behaviors will help me advance within the company.	<i>omitted</i>
I find feedback on my advancement potential useful.	<i>omitted</i>
Feedback-related uncertainty ²⁵	
<i>Alpha 0.92</i>	<i>Alpha: 0.79</i>
When my instructor pilot (IP) gives me feedback about my performance, I'm often confused as to where he is coming from.	<i>omitted</i>
When my IP gives me feedback about my performance, I'm not certain as to what it means in terms of our working relationship.	<i>omitted</i>
When my IP gives me feedback about my performance, I'm usually confused by the feedback.	When my tutors give me feedback about my performance, I'm usually confused by the feedback.
When my IP gives me feedback about my performance, I am not immediately sure what it	When my tutors give me feedback about my performance, I am not immediately sure what it requires me to do.

requires me to do.

When my IP gives me feedback about my performance, I'm not sure how the feedback will figure into my daily grade slip.

omitted

When my IP gives me feedback about my performance, I still really don't know how well I am doing.

When my tutors give me feedback about my performance, I still really don't know how well I am doing.

When my IP gives me feedback about my performance, I'm not certain how best to respond to it.

When my tutors give me feedback about my performance, I'm not certain how best to respond to it.

When my IP gives me feedback about my performance, his motives are not clear.

omitted

When my IP gives me feedback about my performance, I'm not sure what will result from the feedback.

omitted

Frequency of Feedback Monitoring³³

Alpha 0.77

Alpha: 0.76

In order to find out how well you are performing in your present job, how FREQUENTLY do you (5-points ranging from very frequently to very infrequently)

In order to find out how well you are performing as a medical student, how FREQUENTLY do you (5-points ranging from very frequently to very infrequently)

Observe what performance behaviors your boss rewards and use this as feedback on your own performance?

Observe what performance behaviors and skills your tutor rewards and use this as feedback on your own performance?

Compare yourself with peers (persons at your level in the organization)?

omitted

Pay attention to how your boss acts toward you in order to understand how he/she perceives and evaluates your work performance?

omitted

Observe the characteristics of people who are rewarded by your supervisor and use this information?

Observe the characteristics of people who are rewarded by your tutors and use this information?

Frequency of Active Feedback Seeking³³

In order to find out how well you are performing in your job, how FREQUENTLY do you (5-points ranging from very frequently to very infrequently)

In order to find out how well you are performing as a medical student, how FREQUENTLY do you (5-points ranging from very frequently to very infrequently)

Seek information from your co-workers about your work performance?

omitted

Seek feedback from your supervisor about your work performance?

Seek feedback from your tutors about your work performance?

Seek feedback from your supervisor about potential for advancement within the (X) system?

omitted

Chapter 3

Barriers to the uptake and use of feedback in the context of summative assessment

Christopher Harrison
Karen D. Könings
Lambert Schuwirth
Val Wass
Cees van der Vleuten

Published in:
Advances in Health Sciences Education
2015; 20: 229-245

Abstract

Despite calls for feedback to be incorporated in all assessments, a dichotomy exists between formative and summative assessments. When feedback is provided in a summative context, it is not always used effectively by learners. In this study we explored the reasons for this. We conducted individual interviews with 17 students who had recently received web based feedback following a summative assessment. Constant comparative analysis was conducted for recurring themes. The summative assessment culture, with a focus on avoiding failure, was a dominant and negative influence on the use of feedback. Strong emotions were prevalent throughout the period of assessment and feedback, which reinforced the focus on the need to pass, rather than excel. These affective factors were heightened by interactions with others. The influence of prior learning experiences affected expectations about achievement and the need to use feedback. The summative assessment and subsequent feedback appeared disconnected from future clinical workplace learning. Socio-cultural influences and barriers to feedback need to be understood before attempting to provide feedback after all assessments. A move away from the summative assessment culture may be needed in order to maximise the learning potential of assessments.

Introduction

The importance and potential power of feedback is well-established, both in general educational settings and in relation to the health care professions.¹⁻⁴ As a result, there have been calls to provide feedback after all forms of assessment.⁵ However, there often remains an artificial dichotomy between formative and summative assessments. In the former, the focus is on using the assessment to enhance learning. By contrast, the focus in summative assessments is on determining if the examinees are competent to progress to the next level. This often limits the effort teachers are willing to put into providing feedback but also limits the uptake of feedback by students especially those that have passed the assessment.⁴ Though high-stakes assessments have the potential to generate much useful information about the learner, they are often ultimately reduced to a single pass-fail decision. Though this decision-making is clearly important, the lack of feedback provision represents a potential missed opportunity, if it can be achieved without increasing the burden on assessors.

This reductionist approach to assessment has been challenged.⁶ In a high-stakes OSCE assessment, the scores for each item or station are typically added together to make one final mark. Thus, a student who scores full marks on a communication skills OSCE station but zero marks on a resuscitation station will receive the same overall score as one who scores zero marks in communication skills but full marks in resuscitation, even though the information learnt about each learner's competence may be vastly different. Similar analogies apply when totalling up marks from written assessments, such as single-best-answer questions.

Attempts to challenge this artificial dichotomy and instead provide feedback after high-stakes assessments are not necessarily successful. From the limited evidence, it appears that students will not always take up the offer to receive feedback, or will make limited use of it, following summative written assessments or OSCEs.^{7, 8} Performance appears to be an influential factor, with students who perform very well, and those who fail the assessment, tending to make more use of the feedback, whereas those students who just passed make least use.⁸ These studies were quantitative in nature, so cannot explain why students might not engage with feedback in these circumstances. However, if we really want to harness the power of assessment to enhance future learning, we need to understand why learners engage with feedback in a summative context or decide not to. The literature on feedback has been developing rapidly in recent years but rarely considers feedback in a summative context. It is worthwhile considering if any of the conclusions may be applicable to the summative setting.

Although feedback is a highly complex process, there is general agreement that several important steps need to be negotiated in order for it to be successful in improving the knowledge or performance of the learner.^{1, 2, 4} Firstly, a learner needs to be receptive to

receiving it. Secondly, the learner must understand the message being given, so it must be such that it aligns with the learner's frame of reference.⁹ Thirdly, the learner needs to set concrete, meaningful and attainable goals, and then take steps to reach them.^{1, 2} The importance of creating a 'culture of feedback' has been advocated.⁴ Recent work has demonstrated how social and cultural influences within a profession can influence an individual's receptivity and responsiveness to feedback.¹⁰ These issues will be considered briefly for their potential role in affecting learners' receptivity to feedback after summative assessment.

Feedback-seeking behaviour can be defined as "the conscious devotion of effort towards determining the correctness and adequacy of one's behaviours for attaining valued goals".¹¹ Feedback-seeking behaviour by employees in the workplace has been extensively studied by organisational psychologists. Ashford and Cummings¹² proposed three main motives for feedback-seeking behaviour. Firstly, there is the need to obtain useful information; secondly there is the need to preserve or enhance one's own ego; and thirdly there is the need to defend or enhance the impression others hold of the person.¹² For instance, employees, when they are confident that their performance is good, seek feedback not to obtain information for themselves, but rather to convey or highlight information to others.¹³ There is some empirical support that these findings may be applicable to a medical education setting.^{14, 15} Bok et al.¹⁵ found that characteristics of the feedback provider, characteristics of the feedback seeker, and the relationship between the feedback seeker and provider are all important.

In the context of the current study, it may be more relevant to consider why learners may refrain from seeking, or actively avoid, feedback. As well as feedback-seeking behaviour, Moss et al.¹⁶ proposed that active feedback-avoiding behaviour also exists. Poorer performers in the workplace may actively avoid feedback when they are aware that their performance is poor. Other studies have also demonstrated that poorer performers are less likely to seek feedback.^{13, 17} We are not aware of any studies of active feedback-avoiding behaviour in the medical education literature.

Educators need to ensure that students understand the feedback message. Vagueness and uncertainty will lead to a higher cognitive load for learners, which can impair the response to feedback.² Directive feedback, in which clear guidance is given on what needs to be changed and which strengths should be maintained, appears to be most useful for novice or struggling learners, whereas more facilitative feedback, which comprises comments and suggestions, is acceptable for more experienced learners.² At the same time, a very long complex feedback message is not necessarily helpful, as learners may lose interest.²

Even if the feedback message is clearly understood, learners need to be willing to accept the feedback and make use of it. Learners will tend to make little use of feedback if the information provided confirms what they already knew.¹ Feedback may be disregarded if it

is lacking in credibility; for example because the feedback giver has not observed a learner directly.¹⁰ Learners are apprehensive about receiving feedback information which challenges their own self-assessment of their abilities.¹⁸ Critical feedback can induce strong emotional reactions which may block effective use.¹⁹ Some learners appear more interested in using feedback as a means of boosting their self-confidence rather than as a device to correct knowledge or skill deficiencies.²⁰

As well as considering the individual learner's response to feedback, Watling et al.¹⁰ advocate consideration of the social and cultural contexts in which feedback and learning takes place. Socio-cultural theory, which has been used to interpret experiential learning in the clinical workplace, may also help us to gain a greater understanding of the complex learning that occurs, or fails to occur, following summative assessments.²¹ Different professions appear to have different organisational cultures with respect to feedback, which influence how learners perceive the constructiveness and credibility of feedback they receive.¹⁰ This challenges conventional approaches to feedback in medical education, which focus primarily on the feedback delivery skills of faculty members.²²

The effects that assessment can have on learning are clearly important. These have previously been categorised as pre-assessment, post-assessment and pure assessment effects.²³ Studies to date on the learning effects of summative assessment have focussed on the pre-assessment effects and not considered the post-assessment effects.²⁴⁻²⁷ In these studies, there are complex and multifaceted interactions between various factors in an assessment system which together influence learning, sometimes in a positive way, but often in a negative manner. Students pick up implicit messages from the teaching staff and their fellow students regarding the expectations of the learning that is required. Students' perceptions of their ability to bring about a particular outcome, and the 'costs' involved in reaching a desired outcome, may also influence their learning.

Although the literature on feedback is currently expanding rapidly, there is a lack of research into feedback in the context of summative assessment. This study therefore seeks to fill this gap by building on the theoretical models in the educational and organisational psychology literature. We aimed to explore why students failed to make more use of feedback after summative assessment, and which factors were influential.

Methods

Context

The study took place at Keele University School of Medicine in 2011 with students from the third year (out of five). Two-thirds of the way through the year, students have a summative 12-station Objective Structured Clinical Examination (OSCE), including stations on history-taking, clinical examination and practical procedural skills. Students

have to pass at least 8 stations in order to progress to the next year. Each station lasted for 8.5 minutes. New stations are introduced each year to minimise the risk of students 'studying to the test'. Students have to pass the OSCE, as well as written assessments, in order to progress to Year 4. Following the OSCE, they have a further 12 weeks of study within the year, based in general practice and in selected clinical areas of their choice. Individualised feedback about each student's performance was released via a website on the same day the results were published but there was no compulsion for students to view the feedback. Students could choose to look at the feedback in different ways; they could view a breakdown of the skills assessed in each station or across the OSCE as a whole. They could also compare their performance with the cohort as a whole. More details about this have been published elsewhere.⁸ A number of screenshots are shown in Appendix A. In years 1 and 2, the students had received feedback on summative written and practical skills assessments in large group classes. They had not had a formative OSCE earlier in Year 3, but had received more informal teaching regarding preparation for the assessment.

Participants and Sampling

Recruitment of students followed a maximum variation sampling approach to ensure that the sample included students with a wide range of achievement in written and OSCE assessments, and who had used the website in very different ways. Maximum variation sampling helps to achieve information-rich cases for in-depth study.²⁸ To preserve anonymity, and in order to avoid any suggestion of coercion, students received an initial email from an administrator, rather than directly from the researchers. Interviewing continued until theoretical saturation was reached.

Individual, semi-structured, face-to-face interviews were conducted in June and July 2011 by the principal researcher (CH), who had previously had little direct interaction with the students. This method was chosen in order to gain in-depth information and encourage deep exploration of personal matters in a way that is often not possible in a group setting.²⁹ The interviewing technique followed recognised methods of rapport development and use of exploratory open-ended questions recommended by DiCicco-Bloom and Crabtree.²⁹ All the interviews were audio-recorded. The interviews lasted 30 - 60 minutes. Questions explored students' experiences of their recent OSCE, their perceptions of the feedback they received and its impact. The interview protocol is shown in Appendix B. Themes which emerged in earlier interviews were explored in subsequent interviews.

In total, data were derived from 17 participants. It proved hard to recruit students who had only just passed the assessment (passed 8 or 9 OSCE stations). As these were a group who had been shown to have made little use of feedback, recruitment continued to focus on those students who had accessed the feedback in only a limited manner but had a slightly higher level of performance.⁸ Details of the participants are shown in Table 1. Ten (59%) of the interviewees were female, which was identical to the proportion of females in the year group as a whole. Twelve (80% of those who provided information) had entered directly

from school which is comparable to the cohort as a whole (73%). Participants visited the website on one to five separate occasions, viewing between 46 to 316 webpages in 6 – 70 minutes. These figures were closely comparable to the cohort as a whole.

Table 1: Summary of participant characteristics

Student	Gender	School entrant or graduate entrant	Total webpages viewed	Total visits to website	Pages viewed per minute	Total time	OSCE score as %	OSCE stations passed (out of 12)	Quartile rank (based on all assessments)
1	F	G	316	5	4.9	65	81	12	2
2	M	S	311	5	4.9	63	81	12	4
3	F	S	46	1	7.7	6	84	12	1
4	M	S	134	3	6.4	21	82	12	2
5	M	S	111	2	13.9	8	78	10	1
6	M	S	261	3	5	52	75	12	3
7	F	S	93	2	9.3	10	67	8	4
8	F	S	132	2	9.4	14	82	10	3
9	F	S	117	4	1.7	70	70	10	3
10	F	S	279	2	8	35	83	12	1
11	F		75	1	4.7	16	70	11	2
12	F	S	67	1	5.6	12	79	12	2
13	M	S	119	3	7.4	16	81	12	1
14	M	S	232	3	6.4	36	83	11	1
15	M	G	71	2	11.8	6	71	10	2
16	F		64	1	10.7	6	74	11	4
17	F	G	134	2	8.9	15	72	11	3

Notes:

M = male, F = female

S = entered medical school straight from school. G = entered medical school as graduate (two students declined to provide this information)

Quartile rank represents rank in year based on all assessments: 1 is top quartile etc

Data Analysis

Verbatim transcriptions of the interviews were analysed with the assistance of computer-assisted qualitative data analysis software (NVIVO version 9, QSR International Pty Ltd, Doncaster, Victoria, Australia). Data were analysed using the constant comparative method, which comprises simultaneous coding and analysis of data in order to develop and refine concepts and explore their inter-relationship.³⁰ Analysis combined coding which was guided by *a priori* awareness of the feedback literature relevant theories with inductive

coding which emerged from the data. Findings from the interviews were triangulated, with the participants' consent, with information from each participant's usage of the website, and their performance in the assessments.³¹ The principal researcher (CH) coded all transcripts and constructed initial themes. A quarter of the transcripts were reviewed by a second researcher (VW) against the coding framework and themes. A third researcher (SY), unconnected from the main research project, independently coded two transcripts. Discrepancies were resolved by discussion.

Ethical issues

The study was approved by Keele School of Medicine Ethics Committee. Participants were given an information sheet and asked to sign a consent form before the interview commenced. As a token of thanks for their participation, participants were given a gift voucher, worth £10. They were informed of the availability of the gift voucher at the recruitment stage.

Results

Five themes emerged from our analysis: (i) the power of the summative assessment culture and its negative impact on the use of feedback; (ii) the influence of strong emotions; (iii) the influence of social interactions with others; (iv) the influence of prior learning experiences on expectations; (v) the disconnection between assessment and future learning. These key themes will be described in more detail and highlighted by representative comments from participants.

Power of summative assessment culture

Summative assessment created a powerful culture that was dominated by fear of failure and potential punishment, *i.e.* the need to take re-sits.

I always say that I'm happy to just pass. Because it's a pass and I never ever want to have to go through everything again, to resit. (Student 8, Female, passed 10 out of 12 stations, quartile rank 3, spent 14 minutes on the website)

I thought I'd failed it. It was horrible, horrible, horrible. (Student 9, Female, passed 10 stations, quartile rank 3, spent 70 minutes on the website)

The fear of failure, and the need to jump over the assessment hurdle, pervaded most interviews and prevented the OSCE from being regarded as a learning opportunity, but was instead an extrinsic motivator.

Everyone's so worried that they're gonna fail that they just work really hard. (Student 3, Female, passed 12 stations, quartile rank 1, spent 6 minutes on the website)

This culture paradoxically focused more attention on students who failed and required the punishment, while ignoring students who passed the assessment:

It's always, you get your mark and, and if you, if you fail then you see someone about it, and if you pass that's it, you, you go off on your own way sort of thing. Even if the person who passed only got a mark more than the person who's failed. (Student 2, Male, passed 12 stations, quartile rank 4, spent 63 minutes on the website)

The attitude towards the hurdle being seen as an endpoint, not as a stepping stone towards future learning, was perceived as being reinforced by the medical school.

Well I must admit, I think part of the reason why I don't put much thinking into the results might be because, here at the medical school, when the exam's over it's over. They don't mention it, you get on with your normal day-to-day activities." (Student 5, Male, passed 10 stations, quartile rank 1, spent 8 minutes on the website)

Influence of strong emotions

The fear of failure generated very strong emotions in the students. On the day of the OSCE, the wait for the exam to begin was a time of high anxiety, with several participants being physically sick. Fear of failure was dominant while awaiting the results, which led to even greater anxiety. Some students were too fearful to open the results and asked colleagues or friends to do it on their behalf.

The timing of the feedback was a factor. In this instance it was released while they were revising for other written exams and some did not want it to disturb their revision for these exams, as the fear of receiving a failing mark was likened to a grief reaction:

Now, I was in two minds whether to look at it, because, obviously, there was written exams to prepare for.... because I think if I hadn't have done so well, I don't think I would have, er, been in the correct mind frame for my written exams. Because I think whenever you've done bad in exams you need a bit of time to mourn. (Student 16, Female, passed 11 stations, quartile rank 4, spent 6 minutes on the website)

In contrast, the period after the release of results was marked, in the vast majority of students, by an intense feeling of relief, as the goal of avoiding the re-sits had been achieved. This reduced the motivation of students to engage with the specific feedback. The

enormous relief at overcoming the assessment hurdle, following the fear of failure, appeared to block a more rational interest in the feedback. The anxiety, both while performing the OSCE, and while awaiting the results, appeared to intensify the feeling of relief.

And I heard that the results were up. And I was really, really nervous. I had to get my friend to check it for me because I was so nervous. And because of the feelings, like I thought I'd failed and I really didn't want to have to retake them again. But when my friend checked it and I was told that I'd passed, I was over the moon. (Student 8, Female, passed 10 stations, quartile rank 3, spent 14 minutes on the website)

Social interactions with others

Interactions with clinical teachers, peers, parents and partners (boyfriends/girlfriends) reinforced the need to avoid failure. Students perceived that clinical teachers regarded the OSCE as a hoop to jump through, for which a pass was sufficient, whereas the real clinical learning took place separately, and often after graduation.

Some [teachers] say it's about passing and say that the true, erm, medicine begins when you become a doctor. I've heard that quite a few times from clinicians. Some, erm, push for excellence, but don't push for excellence for academic success, but for excellence for your future practice as a doctor. (Student 16, Female, passed 11 stations, quartile rank 4, spent 6 minutes on the website)

There was rarely any in-depth discussion of the feedback between students, other than discussion as to whether they needed to re-sit or not, which reinforced the focus on the summative pass-fail culture rather than on using the feedback for learning. Students perceived a strongly competitive attitude among the cohort, partly due to academic ranking being used in applications for the first jobs when leaving medical school and partly due to a perceived inherent competitiveness from medical students being high academic achievers. This attitude was rarely explicitly discussed between students and they struggled to explain why it existed so strongly. Nevertheless, almost all students sensed the hidden competition and this appeared to influence their response to the feedback. Students often expressed contradictory opinions about their own and others' attitudes to competition:

I find it useful to see how I am doing in comparison to everyone else....I hate it when everyone comes out and asks 'what percent did you get?....I'd like to see myself, but I don't want everyone else to be comparing'. (Student 14, Male, passed 11 stations, quartile rank 1, spent 36 minutes on the website)

Students were reluctant to reveal to others that they had performed well for fear that it would be interpreted by others as arrogant or uncaring towards those who had performed

badly. Others felt unable to keep up with the academically strongest performers, and ‘opted out’ of the competition and aimed just to pass.

I wouldn't say, 'Oh, you know what, I got a good mark,' where some people do, but that's just not me because I just don't like showing off. I don't like people thinking that I think that I'm really good [uh-huh], because there's things that I'm not good at, so I don't wanna be seen as that. (Student 5, Male, passed 10 stations, quartile rank 1, spent 8 minutes on the website)

Students also had to negotiate a complex situation of competition with their partner without upsetting the relationship.

I wouldn't want to make myself sound like I think I'm better than her. And I wouldn't want her to make out that she's better than me, so it works two ways. (Student 5, Male, passed 10 stations, quartile rank 1, spent 8 minutes on the website)

Although students seemed more willing to show the feedback data to parents, this was usually in order to demonstrate that they are performing to a satisfactory level. They valued positive feedback from parents more than from examiners.

Erm, I think, in a way, it was, sort of, like, me showing, 'Oh, look Mum,' like, 'this is the pass rate and I got,' you know, 'three marks higher than the pass rate, that's good, isn't it?' Erm, and that's because I think everyone do seek some, kind of, approval and some congratulations, but that needs to be from someone who knows you on a personal level, and that's why I think I didn't wanna have congratulations from the examiners I didn't know, I wanted it from someone else. (Student 5, Male, passed 10 stations, quartile rank 1, spent 8 minutes on the website)

Parents reinforced the attitude among students to be content with just passing the assessment.

My mum's just like, 'As long as you pass it doesn't matter,' so she's, like, I suppose that's where I've got it from [uh-huh], because when I get stressed, she's, like, 'Oh, it doesn't even matter anyway, just as long as you pass, you're fine.' (Student 14, Male, passed 11 stations, quartile rank 1, spent 36 minutes on the website)

Influence of prior learning experiences on expectations

The culture in which students had learnt prior to attending medical school affected their expectations about achievement and the need to use feedback to improve. Students reported difficulty in adjusting to a change in their expectations for academic achievement; as school

pupils they were used to coming near the top of the class, but now as medical students they were well down the field. This led to a sense of futility at being unable to achieve marks towards the top of the year. As a result, they avoided trying too hard and instead were content with the lower goal of simply passing the assessment. This further reinforced the summative nature of the assessment.

*I think that's probably ... because most people at Med School will get As, A*s, things like that and to get that you get like 90% over and above and marks like that and I think that, I still sometimes compare it in a way. I think that when I did well, in my A-level I you know I got ... I got over 90% in the exam but here I'll get what under 60% and it sort of, try not to compare it but it does feel a bit like well 60% is a lot lower than 90%. (Student 4, Male, passed 12 stations, quartile rank 2, spent 21 minutes on the website)*

Disconnection between assessment and future learning

There was a disconnection between clinical learning in the workplace and the OSCE. While spending extensive time seeing patients in the clinical workplace was seen as essential in order to gain a good mark in the OSCE, feedback from the OSCE was not seen as necessary for future learning in the workplace. It was seen as relevant only for that specific exam (or for a specific question within the exam) and the variability of their performance across stations made it irrelevant to consider their performance of generic skills across several stations. If there were plans to use the feedback in the future, this was in connection with future OSCEs, not workplace learning.

Erm, I know when it comes to OSCEs next year, I'll probably look at the website again and compare my marks, and think, 'Right, okay then, I didn't get as many marks on this bit. I need to work on that bit.' (Student 11 Female, passed 11 stations, quartile rank 2, spent 16 minutes on the website)

Feedback was rarely discussed with tutors in the workplace to encourage further consideration of the feedback after the emotional response to the release of the results had subsided. When it did take place, this was helpful.

My GP [tutor] asked me areas I could improve so....I looked at the OSCE thing to see which stations I could have improved on and we went through them, which was quite useful. (Student 14, Male, passed 11 stations, quartile rank 1, spent 36 minutes on the website)

Discussion

We aimed to explore why students failed to make more use of feedback after summative assessment, and which factors were influential. This study has demonstrated how the dominance of the summative assessment culture influences behaviour, emotions and cognitions regarding feedback. This fosters the reductionist approach to assessment, with the main aim to avoid failure. Students are very relieved to pass and, as they feel they have achieved their goal, they perceive little incentive to address their weaknesses. Motivation is known to be a critical factor in the response to feedback.¹ Other studies have previously shown the strong connection between emotions and academic achievement, but the influence on feedback utilisation has not previously been considered in detail.³² The assessments are disconnected from most other clinical workplace learning in the minds of the students and this is reinforced by the attitudes of clinical teachers and family members.

Our study has also demonstrated that students seek to manage the perceptions of others, particularly their parents, by using the feedback to demonstrate how well they are doing, or to show how difficult the course is. This has some similarities with the organisational psychology literature, in which employees seek feedback not to obtain information, but rather to highlight information to others.^{12, 13} We did not find clear evidence of active feedback-avoiding behaviour.

Although there appeared to be little evidence that students failed to understand the feedback message, there was confirmatory evidence that learners were often not motivated to make use of the feedback. This was partly because the students disconnected the OSCEs from clinical workplace learning and partly because the summative culture encouraged a focus on avoiding failure. As such, they had achieved their goal, so students failed to perceive a 'feedback gap'.

The findings from this study reinforce the notion that learning from feedback cannot be dissociated from the context in which it takes place, as described by situativity theory.³³ This theory proposes that educators need to consider not just the method by which information is given to learners, but instead focus on understanding the situations in which the information will, or will not, be used. The learning environment is as important as the way a learner internally processes information. The theory also proposes that authentic learning activities should result in better learning.³³ The summative, high-stakes assessment culture, with its focus on passing and failing, with concomitant rewards and punishments, along with a disconnection from the authentic clinical workplace learning, is critical. Other important elements of this context include the social interactions with others, whether it be fellow students or the more hierarchical influence of teachers and parents. It is important to understand how this context can limit the transferability of learning following an assessment to the learning desired in the clinical workplace. At the same time, it is important to recognise that the way in which learners respond to feedback is not uniform.

Furthermore, learners are actually not simply passive ‘consumers’ of the learning and assessment environment, but are generally ‘agentic learners’ who actively seek to influence their own learning and try to adapt the environment to be maximally conducive to their learning.³⁴ In this light the stifling effect that summative aspect has on the uptake and use of feedback is even more striking.

This study provides further insights into the complex effects that summative, high-stakes assessment has on receiving and learning from feedback. Our findings show some similarities to the pre-assessment effects of summative assessments.²⁴⁻²⁷ In both cases there are complex and multifaceted interactions between various factors in an assessment system which together influence learning, sometimes in a positive way, but often in a negative manner. Students pick up implicit messages from the teaching staff and their fellow students regarding the expectations of the learning that is required.

Implications for Practice

An appreciation of the relevance of socio-cultural theory to feedback in this context leads to some challenging conclusions. Asking faculty to provide more, or better-quality, feedback will have only limited success. Similarly, encouraging students to reflect on their results and their feedback is likely to have little impact. Instead, the focus should be on the development of a ‘feedback culture’.⁴ Achieving this culture within the medical profession seems particularly problematic when compared with other professional cultures such as sports and music, as feedback is often given in a sporadic rather than planned way, and on the basis of inference rather than directly observed performance.¹⁰ Unlike medicine, music and sport both have a culture in which a tutor or coach has a more clearly defined and developed educational relationship with the learner, which facilitates the provision (and acceptance) of more challenging feedback. Nevertheless, the potential value of feedback requires that we face the challenge and consider potentially radical solutions.

This study’s findings about the negative effects of the dominant summative assessment culture therefore support the calls to move away from a pure focus on assessment of learning and towards a more integrated approach which also incorporates assessment for learning.⁶ One of the previous drivers for this has been a recognition that it is inappropriate to rely on a single assessment point, as this is inevitably flawed.³⁵ To counter this, a programmatic approach to assessment was first proposed some years ago.³⁶ Since that time, quality criteria and guidelines for such programmes have been developed.³⁷ More recently a theory-driven framework for the practical implementation of a programmatic assessment model has been proposed.³⁵ This model is designed for learner-centred programmes which favour deep learning strategies and is designed to have three main purposes: it should support assessment for learning; it should enable high-stakes decisions to be made using robust evidence; it should produce information that enables the curriculum to be improved. This model is theoretically coherent and there is emerging evidence that it can be practically implemented in different settings.^{38, 39} Initial evidence suggests it may well be

beneficial in encouraging the development of self-regulated learning behaviour among learners and a feedback culture, though more evidence is needed.⁴⁰

However there are a number of challenges that threaten the practical implementation of this model. Medical schools are often restricted by university regulations or other legal restrictions which embed a summative assessment culture. Aside from this, there may be faculty resistance to whole-scale changes in the programme of assessment and in particular to the perceived downplaying of the importance of stand-alone summative assessment. This may occur despite evidence that summative assessment can have significantly adverse effects on students' learning.^{25, 26}

Limitations

There are some limitations to our study. It was conducted in a single centre with a single year group and used a single summative assessment method. We do not know if participants would have given more or less attention to feedback if the assessment had not been summative. Interview studies are by their nature limited to considering only the participants' perspectives and reports of how they use the feedback. Only one interviewee was from the 'just passed' group and none had failed the OSCE. We have tried to enhance the credibility of the findings by using multiple data sources to triangulate the findings, linking the interview data with the way each participant used the website and performed in the assessment. The findings seem to resonate with much of the existing feedback literature based on formative assessment, as well as the literature about the preparation for written summative exams. Our context is not unusual – as most undergraduate courses have summative OSCEs – and is therefore applicable to many medical schools. This therefore suggests that the findings may be transferable.

Suggestions for Further Research

In this study we have investigated factors which affect medical students' utilisation of feedback in the context of summative assessment. Further research is required to investigate if these factors are replicated in other settings. Before considering abandoning the concept of summative assessment, we need to explore the acceptability, among both faculty and students, of changing to more programmatic form of assessment, in which the focus is on both assessment for learning and assessment of learning. It is also important to study students' engagement with feedback in medical schools which are implementing a programmatic approach to assessment where a feedback culture has been fostered.

Conclusions

All assessments generate large amounts of data about individual students which can potentially inform and enhance future learning. However, the culture of summative

assessment, with its focus on passing and failing, can act as a disincentive for use of this rich feedback information. Yet it is clear that a separation of summative and formative assessments at the programme level does not work; summative assessments do not provide meaningful feedback and purely formative assessment tends to be taken less seriously in the long run. For an assessment for learning programme therefore, the formative function and summative functions need to be combined in all assessments. Our study has shed light on the factors that have to be manipulated in the assessment programme in order to make this combination work.

Acknowledgements

We are grateful to Sarah Yardley for independently coding some of the interviews, to Adrian Molyneux for setting up the feedback website and to Kirsty Hartley for helping to recruit students for the interviews. Finally, we are most grateful to all the students who took part in the interviews.

References

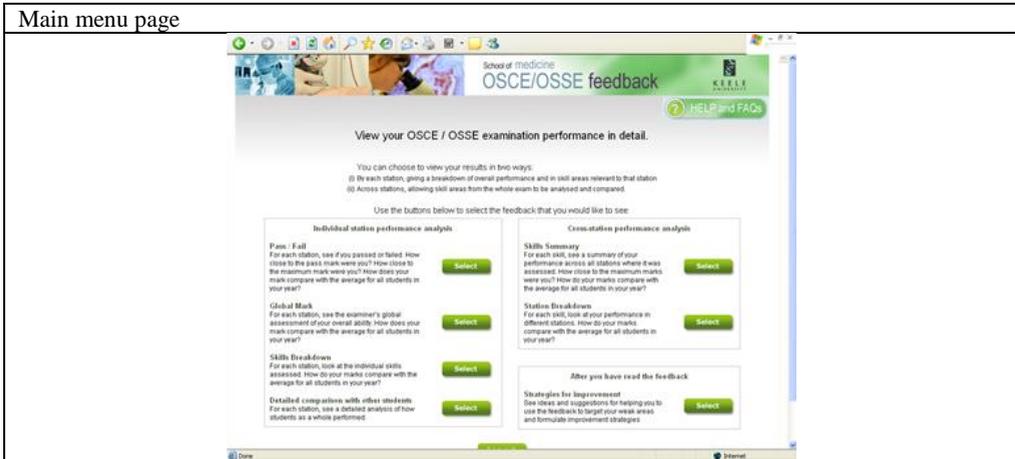
1. Hattie J, Timperley H. The power of feedback. *Review of Educational Research* 2007;77:81-112.
2. Shute, V. (2008). Focus on formative feedback. *Review of Educational Research* 2008; 78: 153-89.
3. Van de Ridder JM, Stokking KM, McGaghie WC, Ten Cate OT. What is feedback in clinical education? *Medical Education* 2008;42:189-97.
4. Archer JC. State of the science in health professional education: effective feedback. *Medical Education* 2010;44:101-8.
5. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher* 2011;33:206-14.
6. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 2011;33:478-85.
7. Sinclair HK, Cleland JA. Undergraduate medical students: who seeks formative feedback?. *Medical Education* 2007;41:580-2.
8. Harrison CJ, Könings KD, Molyneux A, Schuwirth L, Wass V, Van der Vleuten CPM. Web-based feedback after summative assessment: how do students engage? *Medical Education* 2013; 47: 734-744.
9. Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996; 119: 254-284.
10. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education* 2013;47:585-94.
11. Crommelinck M, Anseel F. Understanding and encouraging feedback-seeking behaviour: a literature review. *Medical Education* 2013;47:232-41.
12. Ashford SJ, Cummings LL. Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance* 1983;32:370-98.

13. Morrison EW, Bies RJ. Impression Management in the Feedback-Seeking Process: A Literature review and Research Agenda. *Academy of Management Review* 1991;16:522-41.
14. Teunissen PW, Stapel DA, Van der Vleuten C, Scherpbier A, Boor K, Scheele F. Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Academic Medicine* 2009;84:910-7.
15. Bok HG, Teunissen PW, Spruijt A, Fokkema JP, Van Beukelen P, Jaarsma DA, Van der Vleuten CP. Clarifying students' feedback-seeking behaviour in clinical clerkships. *Medical Education* 2013;47:282-91.
16. Moss SE, Valenzi ER, Taggart W. Are you hiding from your boss? The development of a taxonomy and instrument to assess the feedback management behaviors of good and bad performers. *Journal of Management* 2003;29:487-510.
17. Northcraft GB, Ashford SJ. The preservation of self in everyday life: The effects of performance expectations and feedback context on feedback inquiry. *Organizational Behavior and Human Decision Processes* 1990;47:42-64.
18. Mann K, Van der Vleuten C, Eva K, Armson H, Chesluk B, Dornan T, Holmboe E, Lockyer J, Loney E, Sargeant J. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Academic Medicine* 2011;86:1120-7.
19. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education* 2008;13:275-88.
20. Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17:15-26.
21. Yardley S, Teunissen PW, Dornan T. Experiential learning: AMEE guide No. 63. *Medical Teacher* 2012;34:e102-15.
22. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher* 2012;34:787-91.
23. Dochy F, Segers M, Gijbels D, Struyven K. Assessment Engineering: Breaking down barriers between teaching and learning, and assessment. In: Boud D, Falchikov N (Eds.). *Rethinking assessment in higher education: learning for the longer term*. Oxford: Routledge; 2007. p 87-100.
24. Rudland J, Wilkinson T, Smith-Han K, Thompson-Fawcett M. "You can do it late at night or in the morning. You can do it at home, I did it with my flatmate." The educational impact of an OSCE. *Medical Teacher* 2008;30:206-11.
25. Cilliers FJ, Schuwirth LW, Van der Vleuten CP. A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Medical Education* 2012;12:9.
26. Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, Van der Vleuten CP. The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education* 2010;15:695-715.
27. O'Carroll PJ, Fisher P. Metacognitions, worry and attentional control in predicting OSCE. *Medical Education* 2013; 47: 562-8.
28. Patton MQ. *Qualitative Research and Evaluation Methods*, 3rd edn Sage. Thousand Oaks, California. 2002.
29. DiCicco-Bloom B, Crabtree BF. The qualitative research interview. *Medical Education* 2006;40:314-21.

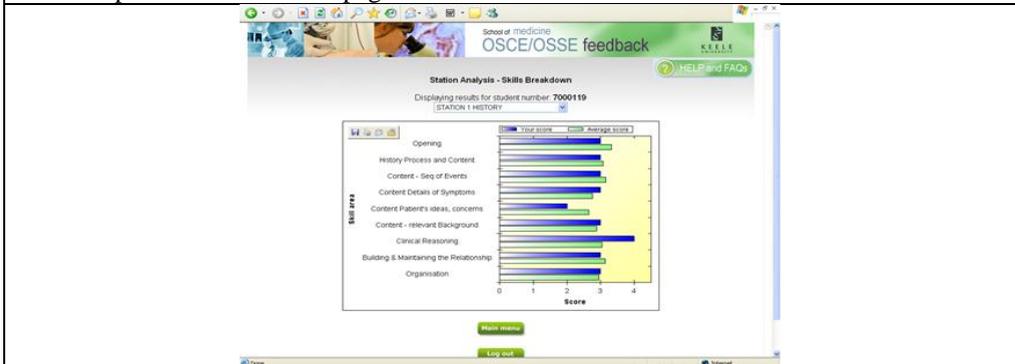
30. Corbin J, Strauss A. *Basics of Qualitative Research*. 3rd ed. Thousand Oaks, CA: Sage Publications; 2008.
31. Lingard L, Albert M, Levinson W. Grounded theory, mixed methods, and action research. *BMJ* 2008;337: 459 – 461.
32. Pekrun R, Goetz T, Titz W, Perry RP. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist* 2002;37:91-105.
33. Durning SJ, Artino AR. Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. *Medical Teacher* 2011;33:188-99.
34. Bandura A. Social cognitive theory: An agentic perspective. *Annual Review of Psychology* 2001;52:1-26.
35. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, Van Tartwijk J. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34:205-14.
36. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Medical Education* 2005;39:309-17.
37. Dijkstra J, Van der Vleuten CP, Schuwirth LW. A new framework for designing programmes of assessment. *Advances in Health Sciences Education* 2010;15:379-93.
38. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007;82:493-502.
39. Driessen EW, Van Tartwijk J, Govaerts M, Teunissen P, Van der Vleuten CP. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Medical Teacher* 2012;34:226-31.
40. Altahawi F, Sisk B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: Experience in a competency-based portfolio system. *Medical Teacher* 2012;34:221-5.

Appendix A: Screenscaptures of webpages

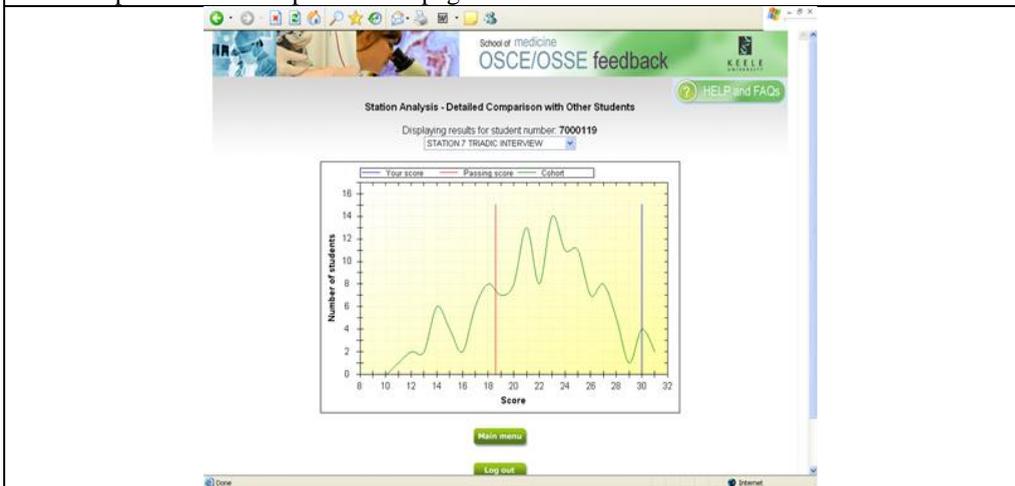
Main menu page



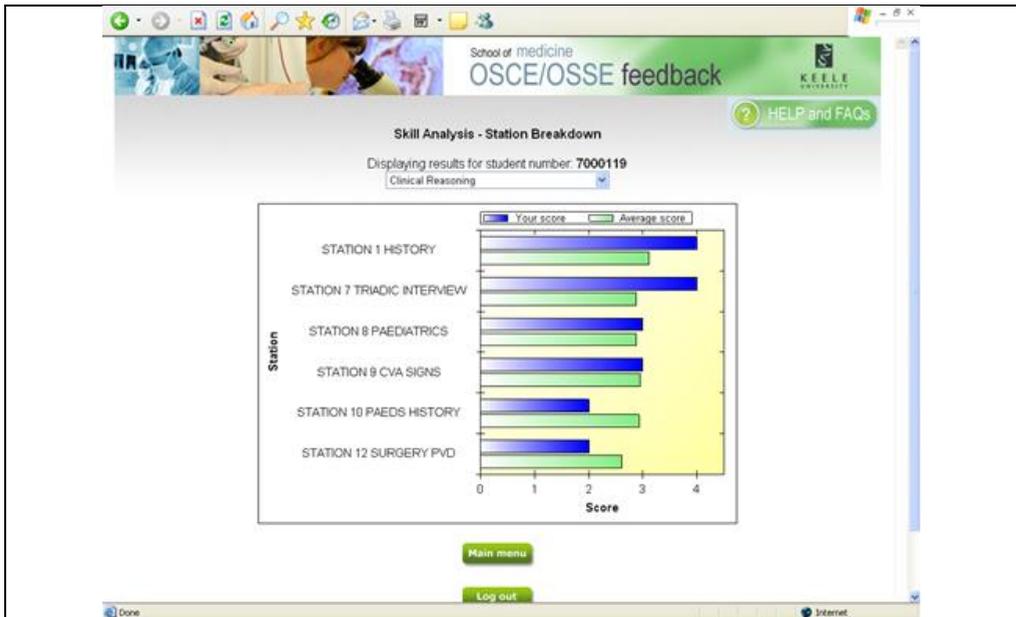
An example skills breakdown page



An example detailed comparison webpage



An example station breakdown webpage



Part of the next steps webpage

Follow-up - Next Steps

What to do after you have read the feedback

We hope that you have found it helpful to read this feedback. We want it to be useful to you in changing your practice. The aim of this section of the website is to enable you to think more deeply about the feedback and enable you to plan for the future.

We suggest you go through these following steps. This is designed purely for your own benefit, so it's important to be honest with yourself. There are a couple of handouts for you to print out.

Step 1
What do you think are your main strengths that this feedback has identified?
 Your performance is likely to be a bit patchy. You will probably have done well in some areas, not so well in others. To start with, focus on the positives. Which skills are you performing well? Write them down on the handout. Don't forget to include all the strengths which the feedback has identified.

Step 2
What are the areas where you need to develop?
 Now turn to the areas where you performed less well. Again, write them all down on the handout.

Step 3
How has this feedback left you feeling?
 It's not always pleasant receiving feedback about yourself. Has it left you feeling upset? Angry? Were the results unexpected? Are you motivated to improve, or are you feeling a bit demoralised at the moment? It's good to acknowledge these feelings to yourself. Again, write the feelings down on the sheet. If the feelings are very strong, you may wish to talk to someone – perhaps a trusted friend, a tutor or a member of the Student Support Team.

Step 4
What are your most important priorities for improvement?
 Now look back at Step 2 and start to think about the main goals you want to work on. You can't solve all the problems at all once as otherwise you may feel overwhelmed. The goals should be:

- **Specific**
 'To be a better communicator' is too vague to be helpful. 'To ask more open questions at the start of a consultation' is a much more specific and useful goal.
- **Measurable**

Appendix B: Outline Framework for Individual Interviews

About OSCE itself

- How did you prepare for the OSCE this year?
- Please talk me through how you were feeling on the day of the OSCE
- Take me through the OSCE – how was that?
- How were you feeling at the end of the OSCE?

About feedback provision

- Please can you talk me through what you did when you went online to look at the feedback
 - What conclusions did you draw from the feedback?
 - How could we make the feedback better for you?
 - Has the feedback made you think about how you would approach a particular station in a different way in future?
 - Do you think the feedback will make a difference to how you will work on the wards, in clinics or in practice?
 - Does it matter to you how well you have done in an OSCE, or is it just important to have passed? Why?
-

Chapter 4

Factors influencing students' receptivity to formative feedback emerging from different assessment cultures

Christopher Harrison
Karen D. Könings
Elaine Dannefer
Lambert Schuwirth
Val Wass
Cees van der Vleuten

Published in:
Perspectives on Medical Education
2016; 5: 276-284

Abstract

Introduction: Feedback after assessment is essential to support the development of optimal performance, but often fails to reach its potential. Although different assessment cultures have been proposed, the impact of these cultures on students' receptivity to feedback is unclear. This study aimed to explore factors which aid or hinder receptivity to feedback.

Methods: Using a constructivist grounded theory approach, the authors conducted six focus groups in three medical schools, in three separate countries, with different institutional approaches to assessment, ranging from a traditional summative assessment structure to a fully implemented programmatic assessment system. The authors analyzed data iteratively, then identified and clarified key themes.

Results: Helpful and counterproductive elements were identified within each school's assessment system. Four principal themes emerged. Receptivity to feedback was enhanced by assessment cultures which promoted students' agency, by the provision of authentic and relevant assessment, and by appropriate scaffolding to aid the interpretation of feedback. Provision of grades and comparative ranking provided a helpful external reference but appeared to hinder the promotion of excellence.

Conclusions: This study has identified important factors emerging from different assessment cultures which, if addressed by programme designers, could enhance the learning potential of feedback following assessments. Students should be enabled to have greater control over assessment and feedback processes, which should be as authentic as possible. Effective long-term mentoring facilitates this process. The trend of curriculum change towards constructivism should now be mirrored in the assessment processes in order to enhance receptivity to feedback.

Introduction

For the development of optimal clinical performance, the importance of linking feedback with deliberate practice is generally acknowledged.¹⁻³ Regulatory authorities have called for feedback to feature strongly in medical training.^{4, 5} Medical students demand more feedback after assessment^{6, 7} and much advice is available for faculty on feedback delivery.⁸ It is therefore surprising that learners sometimes neglect opportunities for feedback after assessment. In one study, 50% of students failed to access feedback on an essay examination,⁹ and in another, students just achieving minimal competence in a summative Objective Structured Clinical Examination (OSCE) made least use of feedback.¹⁰

Feedback may not reach its full potential in practice for several reasons. Faculty find it a complex process and fear being perceived as unkind to learners, as they struggle with conflicting aims of improving learners' future performance and building their confidence.¹¹ Learners fear challenges to their own self-assessments,¹² often wanting feedback to boost their confidence, not correct deficiencies.¹³ Feedback not aligned with a learner's perception may be ignored.^{14, 15} Medicine's learning culture may also limit receptivity. In studies of music, athletics and teacher education, formative critical feedback was expected to push students towards excellence, whereas this was less expected within medicine.¹⁶⁻¹⁹

To maximize learning from feedback, factors influencing its uptake in both high-stakes (summative) and low-stakes assessments should be explored. The summative assessment context has been shown to influence behaviour, emotions and cognitions fostering a reductionist approach which aims mainly to avoid failure and focuses attention on failing students who were 'punished' by resitting assessments.²⁰ In contrast, students who passed, even minimally, were ignored by faculty and felt little incentive to address their weaknesses.

Some medical schools have shifted towards 'assessment for learning' and programmatic assessment with a focus on multiple low-stakes assessments combined with rich, narrative-based feedback throughout the period of study.²¹⁻²³ However, implementing this approach in practice may not be straightforward. Recent studies have demonstrated that students may still regard the low-stakes assessments as summative hurdles to be overcome instead of learning opportunities.^{24, 25} It is therefore unclear how well these modifications to the assessment system promote learning from feedback.

Within these emerging differences in institutional assessment cultures, the impact on students' reception to feedback appears complex. We developed a study to gain further insight. Our research question was: 'what are the factors within medical schools' assessment systems which aid or hinder student receptivity to feedback?' We aimed to answer the question by exploring the experiences of students from three medical schools with different approaches to assessment and feedback.

Method

Context

We purposively selected three medical schools known to have different assessment and feedback systems. When considering which schools to include, we considered published reports as to whether the overall assessment programme was intended to promote ‘assessment for learning’ or ‘assessment of learning’. If the programme was predominantly one of ‘assessment of learning’, we nevertheless wanted an institution that explicitly provided feedback to students as this was a fundamental aspect of our research question. Cleveland Clinic Lerner College of Medicine, USA (School A) has a programmatic approach to assessment with narrative feedback. Keele University School of Medicine, UK (School B) has a mix of formative and summative assessments with both numerical and narrative feedback. The Physician-Clinical Investigator Programme at Maastricht University, Netherlands (School C) has a programmatic approach to assessment with both numerical and narrative feedback. More details of each school’s approach are listed in Table 1 and have been published elsewhere.^{10, 20, 25, 26-29}

Data collection

We chose focus groups in order to seek a range of views and to enable expressed ideas to be developed through interaction between participants. We conducted six focus groups in April to June 2014 (two at each school) to gather the perceptions of the student cohort on their feedback at different points in their studies. A faculty member from the local school, unconnected with the research project, recruited students by email. A convenience sampling approach was taken based on the students’ availability at pre-determined times. Incentives to participate were not offered. To encourage discussion of potentially challenging areas, each group deliberately consisted of a small number of students from a single year. More details of the groups are shown in Table 2. The order in which the groups took place was determined by timetabling constraints: B1, A1, A2, C1, C2, B2. Logistical issues prevented the organization of further groups in order to attempt to reach data saturation on each site. All students from Schools A and B spoke English as a first language; students at School C spoke English as a second language, but all were fluent and had no language difficulties.

The focus groups took place in the respective institutions with a single facilitator (CH) and lasted 60-120 minutes. A semi-structured approach was underpinned by open-ended questions designed to elicit students’ perceptions of their institution’s assessment system, their interpretation and use of feedback and any effect of grades on their aspiration to excellence. Questioning evolved according to the participants’ responses. We recorded and transcribed discussions maintaining student anonymity.

Table 1: Summary characteristics for schools used for research

	School A Cleveland Clinic Lerner College of Medicine, USA	School B Keele University School of Medicine, UK	School C Physician-Clinical Investigator Programme at Maastricht University, Netherlands
Programme Overview	5-year graduate entry	5-year undergraduate/graduate entry	4-year graduate-entry Masters
Students per year	32	130	50
Curriculum	Problem-based learning (PBL)	Mixed PBL with lectures	PBL
Assessment system	Programmatic approach to assessment	Mix of formative and summative assessment	Programmatic approach to assessment
Feedback system	Students receive formative narrative feedback from multiple sources. Grades and numerical scores are not used	Students receive feedback after all summative as well as formative assessments. Mix of numerical and narrative feedback	Mix of numerical and narrative feedback
Portfolio	Students compile a portfolio to interpret, analyze and triangulate the feedback received, with the aim of identifying personal strengths and weaknesses. They then write a reflective essay addressing their progress in meeting competencies, citing feedback as evidence	Students compile a portfolio comprising their personal reflections	Students collect all feedback and other evidence into a portfolio which is used for personal reflections
Mentoring	They meet regularly with a mentor, known as a physician advisor (PA). The PA is responsible for reviewing their formative portfolio	Each student is assigned a personal development tutor who meets them twice a year throughout the whole of the five year programme. This tutor is responsible for determining that the portfolio has been completed satisfactorily	Each student receives support from the same counsellor for all four years; the counsellor is not responsible for assessment decisions
Progression decision	Students are required to compile a summative portfolio, which is used to determine progression. This is assessed by a committee; the PA has no input into progression decisions.	Progression determined solely by satisfactory performance in summative assessments	If a particular assessment demonstrates insufficient knowledge or skill acquisition, the student is required to perform further assessments in order to demonstrate satisfactory competence. The assessment information and feedback in the portfolio is evaluated at the end of the year by an independent portfolio assessment committee and used for the high-stakes promotion decision

Table 2: Participants in each focus group

Focus group	School A		School B		School C	
	1	2	1	2	1	2
Year group and part of the course	1	4	3	4	1	4
	Basic science	Clinical	Early Clinical	Clinical	Basic science	Clinical
Number of students	4	3	5	5	5	5

Data analysis

We used a constructivist grounded theory approach. This approach to qualitative data analysis encompasses the notion that interpretation of the data is co-constructed by both researchers and participants.³⁰ By studying the experiences and perspectives of the participants described in the transcripts, we aimed to identify thematic categories of factors that aid or hinder the uptake of feedback. Consistent with grounded theory, analysis occurred alongside data collection and was able to inform the questioning in the later groups. The lead author (CH) coded all transcripts in order to organize the data and identify key themes and concepts. A second researcher (VW) separately coded three transcripts (one from each school). There was close agreement; minor discrepancies were discussed and could be quickly resolved. At regular intervals, the research team conducted Skype teleconferences to refine the conceptual analysis. As our analysis framework assumes that data are co-constructed by interactions between researchers and participants, we provide the following contextual information: CH, LS, VW are medical doctors with a major involvement in medical education research and development; CvdV, KK have backgrounds in psychology and ED in sociology with a major involvement in medical education research and development.

Ethical approval

We obtained ethical approval from Cleveland Clinic Institutional Review Board (14-355 on 20/3/14), Keele University School of Medicine Ethics Committee (24/2/14) and the Netherlands Association for Medical Education Ethical Review Board (NERB349 on 25/5/14).

Results

Four principal themes emerged: (1) Personal agency; (2) Authenticity and relevance of assessment; (3) Grades and comparative ranking; (4) Scaffolding of feedback. They are described below, with illustrative participant quotes to illustrate the themes.

Personal agency

Factors within the assessment systems across all schools promoted, or hindered, students' personal agency, their capacity to act and make choices using their personal abilities within the constraints and possibilities of their context. Agency was promoted by providing

choice, either within compulsory assessments or by providing optional assessments, which enabled students to demonstrate knowledge acquisition to themselves and their tutors. In contrast, compulsory multiple-choice question papers reduced personal agency by preventing students from fully demonstrating their knowledge; the questions were perceived to focus on specific 'random' facts.

I'm often frustrated with the questions because I think, 'Well, I'm actually good in this field but this one particular question, I don't know.' (School C, Focus Group (FG)1)

Students discounted much of the feedback as irrelevant for future learning or assessments. Feedback methods enabling students to act independently were more conducive to supporting personal agency than those restricting autonomy. For example, online questions and feedback, combined with the ability to revisit these at any time, were preferred to didactic tutor-led feedback sessions delivered in a group setting without provision to revisit later. Providing significant autonomy within the assessment and feedback systems fostered an aspiration towards excellence:

It [the School] really lets you....truly reach and seek that excellence because it gives you that time that you can invest in whatever you deem to be important. (School A, FG1)

By contrast, an assessment system dominated by high-stakes assessments discouraged aspiration to improve:

I think sometimes exams can be limiting because you say, 'Oh, I don't need to know that for an exam.'that stifles learning sometimes. (School B, FG1)

Students felt in control if feedback from low-stakes assessments appeared predictive of future performance in high-stakes assessments. The feeling was the opposite if low-stakes assessments were seen as too dissimilar from the 'real thing'.

Institutional requirements for assessments (e.g. formatting rules for completing a portfolio) limited agency and provoked much frustration and risked devaluing learning. Students sometimes tried to subvert the marking criteria to maximize learning, demonstrating tension between student agency and institutional control.

Instead of looking at what they wanted I just went on a tangent and wrote. And then that for me I was like, 'I know I might not get a good mark. I might not even pass, but this is actually reflective. That's going to be useful for me in the future to look back at.' (School B, FG1)

Similarly, agency was threatened when institutions set criteria for selecting specific items from overall aggregated feedback to demonstrate progress. Students perceived being forced to include critical feedback, at the expense of detailed, personal and positive feedback, even if it was unhelpful, bland and generic. They felt forced to play the ‘assessment game’.

I had feedback, 'You're not saying enough.' But on the other hand I got feedback, 'But the quality of what you're saying is very high.' But I was always forced to make some goals from, 'You're not saying enough,' so I was forced to say more. (School C, FG1)

Students recognized that autonomy should have limits. The safety of future patients necessitated that assessment processes identify students’ weaker areas or flaws in self-assessment. Standardization of assessments was generally accepted, even though it limited autonomy, but acceptance was much lower if it was perceived to lack immediate relevance to their future role as doctors.

If we're going to do a critical appraisal of qualitative research, why not get people to find their own piece? So if you were doing paediatrics at the time, you could have found a paediatric paper that was relevant rather than writing about bottled water because it's not got any clinical relevance to what we're doing. (School B, FG1)

Authenticity and relevance of assessment

Many assessments were seen to lack authenticity or direct relevance to students’ future work as doctors. Feedback was often ignored if assessments were perceived as irrelevant. There was a belief that it was necessary to say and do certain things to pass a clinical skills assessment (such as an OSCE); these were different from what would be done in real-life clinical practice. Feedback given on wards before an OSCE was often ignored and perceived to harm the chances of passing. In contrast, an assessment system which relied only on feedback within the clinical workplace, in the absence of high-stakes OSCEs, avoided giving students mixed messages. It allowed them to concentrate on feedback they received. Recognition that learning, and developing clinical competence, was an evolving process continuing after graduation appeared particularly helpful:

It's kind of comforting that this process [of ward-based feedback] continues as we're residents and that we learn at a similar rate that we're learning here in medical school. (School A, FG2)

Grades and comparative ranking

Grades and comparative student rankings within the cohort had mixed effects on receptivity to feedback. Grades reassured students they were achieving the standards required for qualification:

I think it's like comparing my grades to what they think is necessary. So the necessary level of knowledge to be a doctor, is my knowledge lower than that? Is it higher than that? Is it about right? (School B, FG1)

Grades provided some clarity on the expected level of performance and avoided the potential for narrative feedback to mislead:

Well, I mean, you can get a warm and fuzzy description and be four out of ten. If you get a four then you know you need to do something to get to the six that is required to make the grade. I would much rather know that I need to work harder than have a warm and fuzzy feeling. (School B, FG2)

Information comparing performance with peers was an additional stimulus to take feedback seriously:

If I got a ranking [position relative to peers] of 50 on my MCQs and a ranking of 20 on my KFPs [Key Feature Problems], then I know I need to go and work on my MCQs. (School B, FG1)

Grades following summative, end-of-year assessments gave sufficient information; more detailed feedback was seen as superfluous:

So what was on our Year 2 paper - most of it is irrelevant to what's going to be on our third-year paper. So to give us feedback is pointless. We've got a grade. That's all we need to know. (School B, FG1)

At other times within a programme, grades without narrative feedback failed to provide sufficient information or motivation to stimulate improvement. At one school, so many students were graded 'at the level expected' that it was felt futile to aspire towards excellence.

And for these reports, I don't really feel stimulated to do my very best to get a good grade because I know it will just be on the threshold and I'll be fine. (School C, FG1)

The transition from a grade-based system in a previous programme to one based on narrative feedback without grades presented a challenge for students to then find ways of checking they were 'on track'.

You don't have that oh, I got an A. Okay, you know, I can feel good about myself. It's just...you don't have that external validation. (School A, FG2)

As students adjusted to this new assessment model, they became less dependent on validating labels and focussed more on preparation for practice:

You can see that this is preparing you for real life when you don't get an A every time you go. (School A, FG2)

As the transition continued, the absence of grading or ranking forced students to take a more nuanced view of their peers' strengths and weaknesses. This gave them more freedom in using the feedback while preserving their own self-esteem. The lack of reassurance from 'good enough' grades incentivized students to aim for excellence:

We don't know what the bare minimum is. You better have a slam-dunk to make sure. (School A, FG1)

The tendency for tutors at all schools to provide vague positive feedback frustrated students; it failed to meet aspirations to improve performance.

It means very little to me to always get these 'great job, great job, great job' versus someone who is trying to find ways to help me get better. (School A, FG2)

It was also commonly perceived that tutors feared giving critical feedback. Students at School A described how they had actively challenged tutors for more constructively critical feedback. There was little evidence of feedback-seeking behaviour at the other sites:

They have to be prompted because they're afraid to write something bad in our feedback, but it's necessary. (School A, FG2)

Scaffolding of feedback

Scaffolding, as provided by mentors, played a significant role in assisting, or occasionally hindering, receptivity to feedback. Successful mentoring helped students interpret feedback while allowing them to remain in control. Agency was promoted by reducing the scaffolding as the course progressed to prepare students for a working environment reducing dependency on feedback:

My PA [mentor]... [said]... the whole point of this system is that at some point in your career, you're not going to have people giving you monthly evaluations necessarily,you're supposed to be able to self-identify this and reflect and then do that. (School A, FG2)

A student mentoring system which provided long-term supervision (over several years), and allowed mentors to see the student's entire feedback, enabled trust to develop. Awareness of previous feedback aided interpretation of new information:

They also know all your previous feedback, so they can kind of help you contextualize it. (School A, FG2)

Organizing the support systems in this way enabled mentors to safely challenge students' inaccurate self-assessments:

I tend to be very hard on myself to begin with, and so I would actually meet with my PA, and she goes, ... this is what I'm seeing in the evidence, and this is what you're saying and they don't really match. (School A, FG2)

Scaffolding did not always enhance students' agency. Mentors could become paternalistic and decide what feedback to exclude from the portfolio:

They've got all different opinions about what's good. So then one counsellor comes and says, 'No, you should change this.' And then next you get another counsellor who's also going to check your portfolio and then suddenly it's all wrong and you've got to change it back. (School C, FG2)

Discussion

We aimed to explore factors within different medical school assessment systems affecting student receptivity to feedback and have demonstrated several important factors. Offering students choice and independence within assessment systems promoted receptivity to feedback. Assessments perceived to lack relevance and authenticity hindered openness to feedback. By contrast, assessments and feedback were valued if authentically aligned to their future work. Grades and comparative ranking marks superficially reassured students they were 'good enough'. Not providing grades caused initial uncertainty, but later promoted more authentic recognition that both self and peers had a complex set of strengths and weaknesses which could not easily be labelled. This encouraged an aspiration to excellence. Long-term mentor relationships assisted both students' interpretation of feedback and enabled inaccurate self-assessment to be challenged, as long as mentors avoided paternalistic attitudes which reduced students' agency. Students ignored feedback lacking credibility or quality.

Previous studies have shown the influence of learning culture in modulating receptivity to feedback. In music and sport, a long-term close working relationship with a mentor enables critical feedback to be accepted within an atmosphere of trust.¹⁹ In longitudinal integrated clerkships, authentic assessment and feedback and a supportive mentoring relationship can promote learning.³¹ To our knowledge, the importance of enabling greater student choice and independence within assessment processes to foster greater receptivity to feedback has not previously been described.

Implications for medical education

In recent decades behaviourist approaches to learning, relying on passive knowledge acquisition, have moved towards constructivist approaches and active learning strategies.³² By contrast, the fundamental approach towards assessment has remained behaviourist. Students are rewarded for passing or punished for failing. Many institutions, including those with active learning approaches, have maintained assessments that reward superficial learning strategies, which students unsurprisingly adopt.^{33, 34} Although more constructivist principles for assessment have been requested,³⁵⁻³⁷ implementation has been limited. Our study demonstrates the potential benefits of three key constructivist principles: (i) improving the authenticity of the assessment, (ii) empowering students with a more active role and (iii) gradual descaffolding to enable transformation from a performance orientation towards a learning orientation. Active involvement in assessment and feedback processes is, however, neither easily achieved nor a panacea, and the role of the authenticity of assessments and the credibility of the mentor/coach as a knowledgeable other are important factors in this transformation. There are clearly patient safety risks if students are given excessive choice within assessments too early in the curriculum and the ensuing liberty to ignore uncomfortable feedback. Nevertheless, we argue that promoting and gradually developing a culture of receptivity to feedback will ultimately benefit patient safety.

A long-term mentoring relationship appears essential for feedback provision to stimulate effective learning. It enables trust to develop in a safe environment in which learners can be challenged. Although such relationships are common in other professions, medical education often involves multiple short-term clerkships.^{19, 31} As with the transition to problem-based learning, mentors may struggle to adopt facilitative rather than paternalistic approaches.³⁸

The study was not designed to directly compare the programmatic approach to assessment with the more traditional summative assessment approach. Within each school's assessment system we identified both helpful and counterproductive elements which impacted on receptivity to feedback. The generalization of our findings therefore does not come from stable factors that are supposed to be true in all contexts; rather our findings serve to better understand the driving forces to enable adaptability of this knowledge in the design of assessment in various contexts. Designers of assessment and feedback programmes could benefit from incorporating the more helpful elements in their programmes.

Limitations

Our study has some limitations. Only the participants' perspectives could be interpreted by the authors to co-construct meaning which may mean that certain perspectives may be over-represented and others may be under-represented. We do not claim to have achieved data saturation over the six focus groups. So although we think that the themes emerging from our study are important and make good sense, we are confident that there are other interesting themes that would emerge from future replications of our study. As such the

combination of such studies would serve to complete the whole picture. A further support for the sensibility of our co-constructed conclusions is that they resonate with the existing feedback literature and yet allow a more focused 'lens' on the interactional factors between the learner and their learning context. The decision to sample different year groups was made on good grounds, but the differences between curricula may also have led to under-representation of certain perceptions, so again we want to stress that we do not assume to have been able to provide a complete picture of all themes. To triangulate students' self-reported perceptions, it would be important in future studies to explore the perceptions of mentors and other faculty regarding the students' receptivity to feedback.

Suggestions for further research

Our results suggest several opportunities for further investigation. While it may be clear that there is a need to gradually shift the control over the assessment process from teacher to learner it is important to better understand the factors that could slow down or accelerate this process. Future studies seeking to understand these factors would be highly informative for assessment design. As stated above it would also be important to replicate our study in a different context to establish whether important themes regarding the uptake of feedback would be detected that our study did not find. Finally, research into the conceptualizations of faculty is of paramount importance to allow for triangulation of these themes.

Conclusions

This study has sought to understand the key elements within an assessment system which influence receptivity to feedback. Whether a medical school employs a summative assessment-based structure or adopts the principles of programmatic assessment, we should strive to make the assessments as authentic as possible, consider carefully the benefits and risks of awarding grades and use long-term mentoring to enable students to be both supported and challenged by the feedback they receive. Students perceive that they benefit from a greater degree of control over assessment and feedback processes. The trend of curriculum change towards constructivism should now be mirrored in the assessment processes. Though challenging we believe this can be achieved.

References

1. Shute V. Focus on formative feedback. *Review of Educational Research* 2008;78:153-89.
2. Hattie J, Timperley H. The power of feedback. *Review of Educational Research* 2007;77:81-112.
3. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical Education* 2007;41:1124-30.
4. General Medical Council. *Tomorrow's Doctors*. 2009. Available from: http://www.gmc-uk.org/Tomorrow_s_Doctors_1011.pdf_48905759.pdf

5. Liaison Committee on Medical Education. 2009. Functions and structure of a medical school: LCME accreditation standards. Available from <http://www.lcme.org/publications/functions2013june.pdf>
6. Duffield KE, Spencer JA. A survey of medical students' views about the purpose and fairness of assessment. *Medical Education* 2002;36:879-86.
7. Dahlin M, Joneborg, N, Runeson B. Stress and depression among medical students: a cross-sectional study. *Medical Education* 2005;39:594-604.
8. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher* 2012;34:787-91.
9. Sinclair HK, Cleland, JA. Undergraduate medical students: who seeks formative feedback? *Medical Education* 2007;41:580-2.
10. Harrison CJ, Könings KD, Molyneux A, Schuwirth L, Wass V, Van der Vleuten CPM. Web-based feedback after summative assessment: how do students engage? *Medical Education* 2013;47:734-44.
11. Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education* 2012;46:201-15.
12. Mann K, Van der Vleuten C, Eva K, et al. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Academic Medicine* 2011;86:1120-7.
13. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17:15-26.
14. Sargeant J, Mann K, Sinclair D, et al. Learning in practice: Experiences and perceptions of high-scoring physicians. *Academic Medicine* 2006;81:655-70.
15. Eva KW, Munoz J, Hanson MD, Walsh A, Wakefield J. Which factors, personal or external, most influence students' generation of learning goals? *Academic Medicine* 2010;85:S102-5.
16. Watling CJ. Unfulfilled promise, untapped potential: feedback at the crossroads. *Medical Teacher* 2014;36:692-7.
17. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education* 2013;47:585-94.
18. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Medical Education* 2013;47:842-50.
19. Watling C, Driessen E, Van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Medical Education* 2014;48:713-23.
20. Harrison C, Könings K, Schuwirth L, Wass V, Van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education* 2015;20:229-45.
21. Shepard L. The role of assessment in a learning culture. *Educational Researcher* 2000;29:4-14.
22. Schuwirth LW, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 2011;33:478-85.
23. Van der Vleuten CPM, Schuwirth LW, Driessen E, et al. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34:205-14.
24. Bok HGJ, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education* 2013;13:123.

25. Heeneman S, Oudkerk Pool A, Schuwirth LWT, Van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning – the theory versus practice. *Medical Education* 2015;49:487-98.
26. Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Medical Teacher* 2015;37:323-6.
27. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007;82:493-502.
28. Dannefer E, Bierer B, Gladding SP. Evidence within a portfolio-based assessment program: What do medical students select to document their performance? *Medical Teacher* 2012;34:215–20.
29. Altafawi F, Sisk, B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: experience in a competency-based portfolio system. *Medical Teacher* 2012;34:221-5.
30. Mills J, Bonner A, Francis K. The development of constructivist grounded theory. *International Journal of Qualitative Methods* 2008;5:25-35.
31. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Medical Education* 2013;47:362-74.
32. Dolmans D, de Grave W, Wolfhagen I, Van der Vleuten C. Problem-based learning: future challenges for educational practice and research. *Medical Education* 2005;39:732-41.
33. Savin-Baden M. Understanding the impact of assessment on students in problem-based learning. *Innovations in Education and Teaching International* 2004;41:221-33.
34. Struyven K, Dochy F, Janssens S. Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education* 2005;30:325-41.
35. Stefani LAJ. Assessment in partnership with learners. *Assessment & Evaluation in Higher Education* 1998;23:339-50.
36. Schuwirth LWT, Van der Vleuten CPM. Changing education, changing assessment, changing research. *Medical Education* 2004;38:805-12.
37. Rust C, O'Donovan B, Price M. A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education* 2005;30:231-40.
38. Dornan T, Scherpbier A, King N, Boshuizen H. Clinical teachers and problem-based learning: a phenomenological study. *Medical Education* 2005;39:163-70.

Chapter 5

Changing the culture of assessment: the dominance of the summative assessment paradigm

Christopher Harrison
Karen D. Könings
Lambert Schuwirth
Val Wass
Cees van der Vleuten

Published in:
BMC Medical Education
2017; 17: 73

Abstract

Background. Despite growing evidence of the benefits of including assessment for learning strategies within programmes of assessment, practical implementation of these approaches is often problematical. Organisational culture change is often hindered by personal and collective beliefs which encourage adherence to the existing organisational paradigm. We aimed to explore how these beliefs influenced proposals to redesign a summative assessment culture in order to improve students' use of assessment-related feedback.

Methods. Using the principles of participatory design, a mixed group comprising medical students, clinical teachers and senior faculty members was challenged to develop radical solutions to improve the use of post-assessment feedback. Follow-up interviews were conducted with individual members of the group to explore their personal beliefs about the proposed redesign. Data were analysed using a socio-cultural lens.

Results. Proposed changes were dominated by a shared belief in the primacy of the summative assessment paradigm, which prevented radical redesign solutions from being accepted by group members. Participants' prior assessment experiences strongly influenced proposals for change. As participants had largely only experienced a summative assessment culture, they found it difficult to conceptualise radical change in the assessment culture. Although all group members participated, students were less successful at persuading the group to adopt their ideas. Faculty members and clinical teachers often used indirect techniques to close down discussions. The strength of individual beliefs became more apparent in the follow-up interviews.

Conclusions. Naïve epistemologies and prior personal experiences were influential in the assessment redesign but were usually not expressed explicitly in a group setting, perhaps because of cultural conventions of politeness. In order to successfully implement a change in assessment culture, firmly-held intuitive beliefs about summative assessment will need to be clearly understood as a first step.

Background

Over the last few years, there have been calls within the general educational establishment to move away from the dominance of high-stakes testing at the end of a period of learning (usually referred to as ‘assessment of learning’) towards a focus on multiple low-stakes assessments throughout the period of learning, combined with rich feedback (usually referred to as ‘assessment for learning’).¹ Within medical education, there has been a developing interest in programmatic assessment, which seeks to incorporate both assessment for learning and assessment of learning.^{2,3} Although in some cases this appears to have been implemented successfully, the introduction of this novel approach has been more problematical in other institutions.⁴⁻⁸

Although designers of assessment programmes may be attracted by the theoretical concepts of programmatic assessment, they may be hesitant to proceed with a radical change of assessment culture because of the reported difficulties in implementation. They will also recall that, within medical education, implementation of previous innovations has not always been straightforward. For example, although constructivist approaches to learning, such as problem-based learning (PBL), have gradually become more popular over the last 30 years, the change has often been controversial and not universally accepted by clinical teachers in particular.^{9, 10} Enthusiasts for assessment culture change therefore need to consider the evidence about which factors underpin, or hinder, successful culture change in other fields.

The challenges involved in modifying an organisation’s culture are immense, as organisations are typically inherently resistant to radical change. According to Johnson¹¹, who applies a socio-cultural perspective to organisational change, one of the reasons for this inertia is because the managers responsible for change share common core underlying beliefs and assumptions about the organisational culture. Johnson defines this as a *paradigm*. The assumptions are typically implicit as they are the organisation’s values which are taken for granted. Any potential change tends to be viewed through the filter of the paradigm. This paradigm of beliefs is itself part of a wider ‘cultural web’ of an organisation. Johnson^{11, 12} defines a number of elements which make up the cultural web. The *power structures* of an organisation are often closely linked to the paradigm, as the most powerful members of the organisation are likely to be most closely associated with the core assumptions. The formal and informal *organisational structures* tend to reflect power structures and ensure that certain relationships or structures are particularly valued within the organisation. Organisations typically adopt a number of *rituals and routines* that members follow, often without thinking. *Stories* are often relayed down through an organisation, recalling significant events and people from the organisation’s history. These serve as demonstrations of what the organisation values. Similarly, *symbols* can provide important insights into an organisation’s values; at their simplest, symbols may be logos, but they can also be indicators of power, such as reserved car-parking spaces. Organisations

also have formal *control systems* which monitor progress in certain aspects and therefore ensure where attention is focussed. Overall the interaction between the paradigm and the wider cultural web leads to an assumption that “this is the way things are done around here”.¹¹ If managers are faced with pressure to change, they will typically redesign in a way which is consistent with the prevailing paradigm and the wider cultural web of the organisation. So a change in assessment tool, say from multiple-choice questions to short-answer questions, may be relatively easy to implement as this change does not challenge the underlying paradigm. However, a change from an assessment of learning culture to one based on assessment for learning is much more fundamental. Johnson¹¹ argues that the cultural web needs to be made explicit before more radical change can be contemplated.

Once an organisational culture has been understood, successful change also depends on other factors.¹³ In particular, the consequences for the individual must be considered carefully. It is not possible to change an individual’s behaviour simply by changing the culture.¹³ Instead, a climate of ‘psychological safety’ needs to be created in order to allow individuals to feel personally involved. They need to be able to evaluate for themselves the beliefs and values inherent in the new culture, and to examine the consequences for themselves as an individual.¹³

Educationalists have sometimes tended to act as if they were unaware of this evidence from the change management literature, even though it is likely to be equally challenging to implement change of an educational culture or learning environment as it is to change the culture of a business.¹⁴ The importance of involving students as partners within instructional redesign is often overlooked despite evidence that incorporating students’ perceptions can improve redesign.¹⁵ Instead, students’ input is frequently limited to evaluations about the quality of teachers. Failure to incorporate students’ perceptions when redesigning a learning environment leads to feelings of alienation and disempowerment among students, with resultant adverse effects on motivation.¹⁶ In a previous study on the implementation of programmatic assessment, it was clear that assessments designed to be formative were perceived by the learners to be summative, which resulted in reluctance for the students to take part in the assessments.⁷ These perceptions had not been explored at the implementation planning stage.

Teachers are also important stakeholders to be considered when designing instructional innovations, yet they are often not fully involved in the design process, even though they are expected to enact the redesigned learning environment in practice.¹⁷ A lack of interaction between designers and teachers can leave the latter group unsupported while expected to implement educational innovations in contexts which may be less than ideal.¹⁷ This was evident in the previously-mentioned study of programmatic assessment implementation. Failure to fully involve the clinical supervisors at the design stage left them feeling frustrated as they felt the new system did not value their role as judges.⁷ Top-

down implementation of new assessment practices has failed in the past to bring about a desired change in the behaviour of teachers or students.¹⁸

To aid a much closer collaboration between students, teachers and instructional designers, the Combination-Of-Perspectives (COOP) model has been proposed as a way of visualising the different stakeholders involved.¹⁹ This process of incorporating multiple stakeholders' perceptions when (re)designing a learning environment is usually referred to as participatory design.²⁰ Participatory design does not imply that students should have full control over the design process as they do not necessarily have the expertise to design an environment which is optimal for their learning. Instead it is proposed that they should be partners in a design process that also uses the expertise of teachers and instructional designers. Within the general educational literature, some successes have been reported for the use of a participatory design approach in an undergraduate setting.²¹

It would, however, be simplistic to assume that change can proceed successfully once stakeholders' perceptions are acknowledged. One reason why change is so difficult is that many of the stakeholders' perceptions are much more than mere ideas; instead they represent their own personal epistemologies.²² These are intuitive but firmly-held beliefs which we all hold about the world around us. Any clinician who has tried to convince a patient to adopt a healthier lifestyle will recognise the strength, and often fixed nature, of personal epistemologies, which are frequently in direct conflict with strong medical evidence.²³ Similarly, clinicians themselves often fail to incorporate evidence from research trials into their daily practice because the evidence does not sit comfortably with their underlying beliefs.²⁴ This failure to incorporate evidence can also be applied to assessment in medical education. All students and teachers will have experienced assessment in various forms, both within the school education system and since entering medical school. As a result, they are likely to have formed their own beliefs about how an assessment should be. It may not be easy to change these belief systems, even if they are in conflict with evidence-based research which demonstrates how a modern assessment programme should be designed. The previous studies about receptivity to feedback in different assessment cultures provided some evidence about students' epistemologies regarding current assessment and feedback practices, as well as their perceptions about their teachers' beliefs about these matters.^{25, 26}

If a medical school is seeking to consider a change of assessment culture towards one in which 'assessment for learning' is promoted, a number of factors will therefore need to be considered. Clearly it would be vital to engage all stakeholders, who in this case comprise students, teachers, assessors and senior members of faculty. The institution should seek to understand the shared common beliefs which underpin the assessment culture. In addition, the consequences of the change for individual stakeholders need to be considered by exploring their beliefs about the proposed change. The need to address all these issues together helps to explain how difficult it is to bring about organisational culture change.

Given the difficulty in bringing about organisational culture change, incremental change is often preferred.¹¹ There are a number of steps which would need to occur to convert an assessment programme from one based on ‘assessment of learning’ to one which incorporates ‘assessment for learning’ principles. A key initial step would be to ensure that students make more use of post-assessment feedback. Attempts have been made to combine high-stakes assessments with feedback. Harrison et al. delivered feedback via a website to all students following a summative objective structured clinical examination (OSCE).²⁷ Although almost all students viewed the website, there was considerable variation in how intensively they viewed the feedback. In particular, students who had only just passed the OSCE made least use of the website compared with those who had performed well, or those that had failed the assessment. Further work demonstrated that summative assessments created a powerful culture which was dominated by students’ fear of failure and subsequent punishment.²⁵ Interactions with teachers, peers and others reinforced the need to avoid failure and paradoxically focussed more attention on students who failed rather than on those who had passed the assessments. Furthermore, the feedback provided after the summative assessment was not regarded as relevant for future learning in the clinical workplace, but was only seen as relevant for future summative assessments. In a more recent study, students made more use of feedback after a progress test when the test was integrated into a comprehensive program of assessment, which included mentoring with scaffolding of feedback, than when it was used in a summative manner.²⁸

We therefore wanted to explore an institution’s readiness to adopt initial changes which would help an organisation move towards an assessment for learning culture. We were interested in the design solutions which would be proposed by stakeholders, in the factors which would support the implementation of this change and the barriers which would be faced, and the influence of stakeholders’ epistemological beliefs about assessment on any proposed redesign. This led to the specific research questions:

1. When stakeholders are asked to redesign a summative assessment culture to ensure that students make more use of post-assessment feedback, what solutions do they suggest?
2. How do the stakeholders’ personal and collective epistemological beliefs about assessment influence their redesign of the assessment culture?

Method

Context

The study took place at Keele University School of Medicine, which has approximately 130 students per year. It is one of the newest British medical schools, graduating its first cohort of doctors in 2012. Progression through, and exit from, the course relied on a series of

summative assessments. There is a focus on the provision of feedback to all students following summative assessments, although there is no obligation for students to make use of the feedback. The school has a tradition of respecting the student voice in a number of ways. Student representatives serve on all the major school committees (including for assessments). They also provide feedback about the course on a regular basis, via online surveys and face-to-face interactive group meetings with faculty members. At these meetings, students learn what has changed in the course as a result of their feedback. They also make suggestions for further changes.

Feedback is provided following both written assessments and Objective Structured Clinical Examinations (OSCEs). Written assessments comprise a mix of single-best-answer and short-answer format questions. These are aligned to the particular year's learning objectives; there is no longitudinal progress test. For the single-best-answer questions, students receive feedback around their performance in different specialty areas (e.g. respiratory medicine), regarding the specific clinical presentation (e.g. cough) and the focus of the question (e.g. choice of investigation). This information is available on a website. Students can view how their performance compares with the cohort as a whole. For students who achieve a satisfactory mark in the short-answer questions, small group sessions take place to enable students to view their marked papers alongside the 'model' answers. Students with an unsatisfactory performance receive one-to-one feedback. OSCE examiners provide feedback in two ways in Years 3 to 5 of the programme. Firstly, the breakdown of marks for each station, and for each domain across stations, is provided on a website. More details of this have been published elsewhere.²⁷ Secondly, examiners provide verbal feedback of up to 90 seconds by speaking into an iPad once the student has left the room. This is recorded by an app and is uploaded into the feedback website, where it is available for the student to listen to. More details of this have been published elsewhere.²⁹

Participatory Redesign Meeting

We conducted a small-group meeting based on the principles of participatory design. Participants were purposively sampled to recruit five medical students (from Years 3 to 5 of the programme), two clinical teachers (with responsibility for Years 3 to 5) who are also experienced OSCE assessors and two senior members of the faculty, who have responsibility for overall curriculum design (and who were known to already have an awareness of the broad concepts of 'assessment for learning'). By aiming to have more than half the participants as students, it was hoped to minimise hierarchical power issues within the group. Potential students were approached by members of the School's academic and administrative team (not involved in the research) solely on the basis of their likely confidence to cope in a meeting with senior members of faculty. Previous involvement in faculty structures or curriculum development was not considered in the recruitment process.

A joint meeting was held to include all participants. The lead researcher provided a short presentation explaining why the educational research evidence would support a change in

assessment processes to ensure that feedback was used, along with evidence demonstrating the difficulties in ensuring that students make use of the feedback. The problem with feedback following OSCEs was used as a concrete example, but the participants were encouraged to consider the wider problem of feedback following assessments as a whole. They were challenged to use blue sky thinking to design radical solutions for this problem. Participants spent some time alone writing their own ideas on Post-it notes. They then formed pairs of a clinician or faculty member and student (or in one case a trio comprising two students and a clinician) to share ideas and develop agreed priorities. Participants then formed mixed groups of 4 or 5 to further develop ideas and priorities for change. The group then reformed as a whole to compare proposed solutions and discuss them in more detail. Once they had come up with their suggestions, the group members were then encouraged to critically challenge their proposals in order to explore the underlying factors which would support (or hinder) successful practical implementation of the assessment change. The meeting lasted approximately two and a half hours.

Individual Follow-up Interviews

A potential risk of group meetings is that respondents may provide socially desirable responses, while the strength of their own personal beliefs may not be sufficiently clear. When sharing a room with senior faculty, students and potentially even clinical teachers may not feel sufficiently empowered to voice their concerns about how the change will affect them. As a key element of successful change management requires an understanding of the consequences for individual stakeholders, it was felt important to ensure their beliefs were captured accurately.

Each group participant was therefore invited to an individual follow-up meeting in the following weeks. Seven participants were interviewed within one week of the participatory redesign meeting. For logistical reasons, one faculty member was interviewed 8 weeks later and one student was unavailable. At these interviews, participants were reminded of the suggested solutions which the group proposed. Using a minimally-structured interview process, participants were encouraged to consider the personal consequences for them as an individual, if the proposed changes were to be implemented. This approach was designed to allow further exploration of stakeholders' beliefs towards assessment and feedback. The interviews each lasted approximately thirty minutes.

Data Collection

The post-its completed by individuals and pairs were collected and transcribed. Those completed by individuals were labelled as being written by students or clinicians (to incorporate both faculty members and clinical teachers). This was done to achieve sufficient anonymization. The outputs from the sub-groups were written on flip charts to facilitate sharing with the whole group. These were also transcribed. The participatory redesign meeting as a whole, and the individual meetings, were audio-recorded and

transcribed. These were supplemented by field notes collected by an experienced observer, who had a specific remit to observe interactions between group members.

Data Analysis and Interpretation

Data were analysed from a sociocultural perspective, using Johnson's cultural web as a lens, in order to understand aspects of the organisational culture as well as individual beliefs.¹¹ Although the proposed changes to the assessment culture were interesting in themselves, the interaction between participants, and in particular the relative influence of students in moulding the proposals, was of specific interest. The transcripts and field notes from the group meeting were analysed carefully to understand how the proposed changes were developed (or blocked) within the group. We aimed to identify collective and personal epistemologies held by the group as a whole and by individual members.

Coding of all data was performed by the lead researcher. In order to validate emerging concepts, the participatory design meeting transcript was also thoroughly read by a second researcher, while the individual interview transcripts were analysed by a third researcher. Regular videoconference discussions with the wider research team helped modify the coding and the conceptual analysis arising from it. In recognition that our analysis framework assumes that data are co-constructed by interactions between researchers and participants, we provide the following contextual information: CH, LS, VW are medical doctors with a major involvement in medical education research and development; CvdV, KK have backgrounds in psychology. All members have a strong interest in the effects of assessment on learning. The lead researcher knew well the two faculty members who participated and had previously briefly met the clinician participants. He had only met one of the students prior to the group meeting.

Ethical approval was obtained from Keele University School of Medicine Ethics Committee.

Results

Proposals to Enable Greater Use of Post-assessment Feedback

A summary of the suggestions which emerged from the different phases of the redesign meeting is shown in Table 1. Three principal ideas were discussed in detail in the whole group: the need for more authentic assessment, the potential to give feedback without (or before) the issuing of grades and the role of one-to-one mentoring to support the interpretation of the feedback.

Table 1: Development of main ideas through participatory redesign phase and individual interviews

		Need for more authentic assessment	Provision of feedback without (or before) grades	Provision of one-to-one mentoring following assessments	Improved quality of feedback	Feedback regarding correct 'model' answers	Improved multi-source feedback by peers	Other
Individual post-it suggestions	Students	0	1	2	3	4	0	6
	Clinicians	1	0	5	3	1	0	3
Pairs/trios Sub-groups		1	1	4	2	0	0	0
		1	1	2	2	0	0	0
Whole group		Discussed in detail. Largely supported in the group.	Discussed in some detail. Discussion led by two students. Clinicians implicitly critical.	Discussed in detail, but largely dismissed as logistically impractical.	Discussed very briefly – difficult to provide feedback when considering mark of borderline candidates.	Not discussed.	Brought up briefly by clinician near the end of the meeting	Not discussed
Individual follow-up interviews	Students	All discuss need for greater authenticity	One student remains strong advocate, others critical.	All supportive, while recognising practical constraints	Not discussed		Mixed views	
	Clinicians	¾ discuss need for greater authenticity	All are critical	One clinician supportive, others critical	Not discussed		Mixed views	

Authenticity

There was broad agreement that current assessments, especially OSCEs, were often inauthentic and failed to detect the attributes that will be needed for a medical career:

We still go for the sort of high stakes, set piece drama and it does feel like a drama; whether it's the theatricality of the OSCE or the crescendo of stress prior to the written exam. So it's a drama and there's a little bit of value in that in terms of testing resilience and ability to cope with stress but on the whole, stress that we want graduates to be able to cope with is more sustained. Faculty member 2, individual interview

There was widespread resentment from students that engagement with clinical workplace learning failed to be reflected in assessment performances:

I think definitely you really do notice the people who don't go to wards and then you talk to them after exams and they did better in the OSCEs and you think, 'How did that work? I slaved all year.' Student 3, group meeting

The perception was that students were able to do last-minute superficial learning which was rewarded by the assessment processes but which failed to equip them with skills for ongoing workplace learning:

They did two weeks of really good OSCE revision, nailed the OSCEs, come back in three weeks' time and they don't know what's going on. Student 5, group meeting

Time constraints left students frustrated that they were unable to adequately demonstrate their abilities:

I don't think you ever quite leave an OSCE with the time constraint thinking you've really flourished in that exam and you've really shown them, 'This is everything I know on this. If I've forgotten anything then it's fine,' because you're always time constrained. So you're cutting out things in your own head, trying to leave things out just to get it into that little tiny window. Student 4, group meeting

Restrictions on time also frustrated examiners' efforts to differentiate between students:

One of the things that I see a lot of in OSCE stations is that technique is wonderful but they're just doing this stuff in the correct order as they've been taught but not being able to actually produce what the abnormalities are and so they can't then make a proper working diagnosis. Those students who I know go to the wards a lot are able to often come out with the correct diagnosis but we don't have time to tease that out. Clinical teacher 2, group meeting

The scenarios that could be assessed in an OSCE format failed to reflect the realities of the clinical workplace; attempts to address this were clumsy:

In OSCE stations it tends to be you on your own whereas most of the work I've ever done is being part of a team. And that doesn't seem to be assessed very widely in OSCE stations, the team working. Clinical teacher 2, group meeting

The lack of authenticity led to feedback from OSCEs being discounted:

When you get feedback from a short OSCE you're also thinking in the back of your mind, 'Well, if they would have seen me do it on the ward,' or, 'I've done 40 cannulas but they've only seen me do that one and I couldn't get it into the vein on that funny arm but I can do it on the endoscopy list.' Student 5, group meeting

However, there was no consensus about how to make the assessments more authentic. Attempts to redesign were typically minor rather than radical:

Do you think OSCEs could be better designed within the sort of constraints that we have to pick out people who are good clinically? Faculty Member 1, group meeting

Suggestions considered included lengthening OSCE stations, a return to long cases at the end of each clinical attachments, or a desire to make workplace-based assessments more rigorous and 'objective'. Increasing the time for OSCE stations was thought to support better integration of knowledge and skills assessment.

If you had more time as well it would allow more questioning which would allow the examiner to probe more about why you're doing certain things. Student 1, group meeting

It was also felt likely to support pre-assessment workplace-based learning:

You'd have to be pretty gutsy to work up to a half-hour exam not having been on the wards for a while before that. Student 5, group meeting

However, there was an unresolved tension between the need for greater authenticity and the belief in the need to maintain or enhance standardisation and reliability to ensure fairness:

I guess the long cases are much harder to standardise aren't they, I presume that's why we've moved away from them and you can't so easily do a number of long cases so that you kind of get that spread and you get a truer picture. Whereas you can do lots of OSCE stations. But what they're actually going to be doing when they hit the wards is see people for the first time, taking a full history, doing a full examination, trying to come up with something as a result of it aren't they. So in that sense a long case is more reflective of what they're going to do, they're not going to be asked to just go in and do one tiny little bit of somebody for eight minutes. So I could see the point of that but I just, you'd have to work really hard at how you did that in a way that was fair and equitable. Clinical teacher 1, individual interview

Current attempts to square this circle were not seen as successful:

Some of the [workplace-based] assessments you get really aren't up to it - they're not standardised enough. There's too many variables. And I think they should be more formalised. Clinical teacher 2, group meeting

There were also apparent dangers in abandoning current assessment formats as they were seen as helpful preparation for examinations later in the learners' careers:

That sort of OSCE format then goes on into future postgraduate exams and I think because it is such a stressful format I think there are really quite big advantages to having had quite a bit of experience of it so you're not just thrown into it when you've paid a large amount of money to do this exam and a lot rests on it. Clinical teacher 1, individual interview

Feedback without Grades

A single student's suggestion of providing feedback without grades generated significant group discussion. It was advocated as a way of enhancing intrinsic motivation:

The issue is trying to get students more intrinsically motivated rather than just, 'I passed and am competent,' because that's why I would assume the excellent students are looking over their feedback more than the students that have just passed because they are wanting to better themselves. It takes the competency out of the equation and pushes people to progress their performance each time they come to it. Student 1, group meeting

However, many felt that this would result in them converting the narrative feedback into a numerical mark, in order to determine whether or not they had passed the assessment, with adverse effects on their receptivity to feedback:

You'd just be preoccupied with going, 'Have I done enough to pass this, then?' rather than actually taking on board the feedback. Student 2, group meeting

This approach could be potentially misleading as there were mismatches between the narrative feedback and numerical marks:

In our group students were saying that with the audio feedback they're currently getting from the OSCE stations what they tend to find is that the stations that they've done well in there's usually quite a lot of feedback on what they could do to improve it whereas the stations that they've just passed there tends to be quite encouraging feedback and not so much of the what they could do better. So I don't know how well you can link it to the grade, really. Clinical teacher 1, group meeting

The mismatch was exacerbated by a focus on the pass-fail mark, especially for those students whose performance was only just satisfactory. Examiners needed to focus primarily on the mark, which was seen as the more important issue, with feedback of secondary priority.

I give a lot of feedback, probably too much feedback, to the ones at the brilliant end with all the things they can do better because you don't have to fret about it. The ones who are borderline, you're fretting about them and then you've only got a

few seconds to give the feedback and so it messes up the feedback to those who particularly need it. Clinical teacher 2, group meeting

Failure to award grades was seen as incompatible with one of the fundamental purposes of assessments:

Obviously at some point there has to be a decision of whether you're fit to practise or not so there has to be some sort of a grade at least at the end. Clinical teacher 1, individual interview

Grades provided clarity and reassurance:

I think they give you some security in knowing that you're doing well and where you're at. And I think grades are a lot easier to process in student's mind, say if it was A, B, C or whatever. If you were looking at a C grade you know you'd need to probably up your game whereas if you're at an A, what you're doing is right. Student 4, individual interview

More fundamentally, there was a lack of belief in the evidence underpinning the provision of feedback without grades:

This is a difficult one because I know what I'm meant to think about it, because I know that there's reasonably good evidence that feedback is more effective without the grades but I can't quite believe it. Faculty member 2, individual interview

Mentoring

There was much discussion about the theoretical benefit of some form of mentoring. The discussion concentrated mostly on one-to-one post-exam feedback (how to do better in future exams) rather than long-term supportive coaching. Mentoring was seen as a way of encouraging students to engage meaningfully with the feedback.

You all get bad feedback sometimes and it can be a bit of, it's hard to see the wood for the trees kind of thing if you've got quite a lot of things that you need to kind of like improve upon and I think talking to someone can first of all make you think about it more rationally, because if you get bad feedback it can just seem a bit oppressive. Student 1, individual interview

Many barriers to successful implementation were identified. The timing of any mentoring needed to be soon after receiving post-assessment feedback, but also needed to be long enough to be meaningful. Students recognised that clinical service commitments meant this would be unlikely to be achievable.

I think that's [mentoring] probably the least practical thing that we discussed. Probably one of the most beneficial, though, because it would be amazing to do it. But I think that's a very tough thing to do, especially getting specialists to give you feedback on that area because everyone's got so much work on their plate as it is. Student 4, individual interview

A suitable mentor was seen as someone who was familiar with the requirements for the assessments, not necessarily one with an ongoing relationship with the student.

No they don't have to know us at all they just have to know the exam system and be rounded enough to be familiar with all areas of the exam possibly. Student 3, individual interview

A tension existed between having a specialist who could give credible feedback in their own area versus a tutor who could help more broadly with all aspects of assessment. On the one hand feedback from a specialist was seen as more valuable:

I think the student would respect the feedback more that it's come from a cardiologist because I was rubbish at cardiac exam rather than it coming from some other specialty where you go, 'How many have you done? You probably haven't done one of those since med school.' Student 4, individual interview

In contrast, feedback from specialists could impair preparation for assessments:

If you've got a consultant in that aspect, they're going to start throwing extra stuff into that exam that you don't need in an OSCE. So you can almost be, on the flipside, disadvantaged by having a super specialist who's like, 'Oh make sure you check for this, you've done this, you've done that,' because they don't know what you need in an OSCE. Student 2, group meeting

The challenges involved in finding enough mentors to spend enough time with students needed a huge change in organisational culture:

I think in terms of overcoming it, it would mean a fundamental shift in the culture of a medical school faculty and in the priorities that faculty had. Faculty member 2, group meeting

While there was recognition that such a change would be popular, it was not thought to be worth the culture change if the main aim was to improve assessment performance rather than clinical practice:

If it's just so that you can do better in exams I think that has to be questioned. As a student I'd want that but, 'Is that the best way to use resources?' But if actually you perform better in clinical practice as a result of that feedback then the

argument for me would be compelling. I've no idea how we could judge that. Faculty member 1, group meeting

Compromises considered making the mentoring available on a voluntary basis or targeting certain year groups:

Ideally obviously everyone would have it, it would be compulsory but I think as a compromise even just making some appointments available and seeing the popularity and engaging in it because at the moment obviously you're only getting those appointments if you're failing. Student 2, individual interview

Having thought about it a bit, the logistics of doing it for all five years, I think, would make it very hard but we may be able to identify key points in the five years, where we can deploy the people who would be good at it. It may be that that each student has this sort of experience twice during their five years. Faculty member 2, individual interview

The consensus was that, while popular, this change could not easily be practically implemented:

The conclusions of everybody having a one to one, making sure everybody engages with it would be amazing. And just briefly talking to my housemates about it they're like, yeah that would be so good because I think there'd be a lot of support for it but the blue sky thinking isn't reality. Student 2, individual interview

The Influence of Personal and Collective Beliefs on Assessment Redesign

Within both the group discussion and the individual interviews, the summative paradigm was a dominant factor. Participants' prior experiences of assessment influenced redesign proposals. Within the group, there was evidence of a hierarchy with senior members exerting more influence.

The summative paradigm was dominant

Most of the beliefs expressed were firmly rooted within the summative assessment paradigm. Discussions were dominated by the need to get through assessment hurdles, rather than becoming a good doctor. The primary focus of assessment was the pass-fail mark and the need to prevent unsafe students qualifying as doctors:

So the good and excellent, that's almost an irrelevance. It's the ones that are no good that we're trying to find, surely. Clinical teacher 2, group meeting

The paradigm included a belief in the need for numbers/grades as a form of perceived objectivity or rigour. There were negative attitudes to changes to assessment models if this rigour and objectivity were to be lost. Within this paradigm, the aim of the feedback was

principally that it should help the student do better in future assessments, or allow more chance to check on the accuracy of the marking.

We should have a one-to-one meeting pretty much after the exam and one before the next assessment to see, 'What's the action plan to do well and where can you improve from last time?' Student 5, group meeting

There was a sense that for some, this paradigm worked for them and the majority of students, so they were reluctant to change it radically.

Influence of Prior Assessment Experience

Prior experience of assessment and feedback, whether positive or negative, acted as another filter through which proposed changes were viewed. Participants frequently recited stories of their own personal assessments. As these typically occurred within a summative assessment framework, they acted as a further reinforcement of the summative paradigm. For example, discussion about the need for greater authenticity in assessments reverted to the need for standardisation:

When we had real patients this year I think some people get different findings you think, 'Oh Christ. Was it all the same?' Student 4, group discussion

However, prior experience of a different assessment culture supported desire for change. While most participants felt mentoring was impractical, one member was a strong advocate because of the impact on his own career:

From my own personal experience, when I was at Medical School, I was sadly ignored for several years and then an inspirational mentor came along, and then I grasped what it was all about and it drove me to improvement. Clinical teacher 2, individual interview

Power was persuasive within the group

Although superficially there was good interaction between group members, with students contributing frequently and in detail, they appeared less able to influence the group's opinion as a whole. In contrast, a single clinician's suggestion about the need for greater authenticity appeared to 'strike a chord' with most of the rest of the group. In the end, virtually every member seemed broadly in agreement that assessments are often inauthentic, with students particularly vocal in support.

Within the group setting, a variety of techniques were deployed by participants to encourage or discourage discussion of proposed changes to the assessment culture. Students commonly provided explicit support to other students expressing ideas, but this behaviour was rarely used by clinicians, who tended to express more implicit support. When disagreeing with other group members, students tended to disagree explicitly with other

students but rarely with clinicians. On the other hand clinicians and faculty members rarely expressed explicit disagreement, especially towards students, preferring a more implicit approach instead, often combined with a questioning style or “sitting on the fence”:

It sounds almost as if the grade gets in the way of learning from the feedback either because of the phenomenon of just passing, so knowledge of the grade changes it, but also somebody like me if I got the feedback without the grade I'd be trying to calculate my grade and so wouldn't necessarily engage. So does the grade get in the way? Faculty member 2, group discussion

The clinicians and faculty members were more explicit in expressing their beliefs in the individual interviews.

I can't imagine getting beyond trying to translate the feedback into a grade. My positive responses in terms of learning would be stymied by my internal algorithms trying to work out whether these words mean I've passed or failed. Faculty member 2, group discussion

Students who remained silent in the group when certain ideas were discussed were much more forthright in expressing their opinion in the follow-up interviews:

Not giving the grades first ... I think mentally that would drive you insane. Student 3, individual interview

Clinicians used a couple of techniques to control the flow of discussion within the group. Firstly, they sometimes abruptly changed the topic being discussed. This was usually successful in ending the group's discussion of a proposed change. Another technique employed was to ask the group a focussed question, which was also effective in moving the group discussion in a different direction. For example, in the middle of a discussion on making assessments more authentic, a clinical teacher said:

Can I ask a question? Do you think that you have too few, the right amount or too many assessments as you're going through your medical school? Clinical teacher 2, group discussion

While the expression of individual beliefs was much clearer in the follow-up interviews, there was little evidence that the beliefs had been changed by what they had heard in the group setting.

Discussion

We aimed to explore how a mixed group of stakeholders would redesign a summative assessment culture to ensure that students would make use of post-assessment feedback. In particular, we were interested in how participants' personal and collective beliefs about assessment influenced the redesign. We found that participants shared common assumptions and beliefs about the importance of the summative assessment paradigm. Discussion about the redesign focussed on the use of feedback to help students pass future assessments, rather than using the feedback to help students become better doctors. Elements of an 'assessment for learning' culture, such as long-term mentoring and the provision of feedback without grades, were considered but not seen as practical ideas for implementation. Participants relied heavily on their own prior assessment and learning experiences to guide their views on what changes were possible or desirable. Although discussion between participants from different backgrounds demonstrated good participation in the group discussion, disagreement was often voiced in an implicit manner and senior clinicians and faculty members appeared to exert more influence than students. Follow up interviews demonstrated that underlying personal beliefs were largely unchanged by the group discussion. There appeared to be a shared common assumption that (to paraphrase Johnson¹¹) "summative assessment is the way things are done around here".

The dominance of the summative assessment paradigm, and the desire to make small changes without disrupting the paradigm, is consistent with Johnson's work on the conceptual challenges organisations face when confronted with evidence of the need to change.¹¹ Johnson argues that an organisation's strategy is based on common, often unspoken, assumptions which are shared by members of the organisation. As a result, elements of an organisation's culture are 'taken for granted'. In our study, participants appeared to take for granted that summative assessment is the way in which students should be assessed. Suggestions for change were therefore predominantly aligned with the paradigm of summative assessment.

Despite contributions from all members, it is unsurprising that the senior clinicians and faculty members appeared more capable of influencing the outcome of the discussion. Members of a medical team formed on an ad hoc basis typically have preconceptions regarding the distribution of power, based on prior experiences or stereotypes.³⁰ The implicit ways in which participants disagreed with each other in the group has similarities with Brown and Levinson's theory of politeness, as described in the field of linguistic pragmatics.³¹ In their theory, they describe how speakers employ strategies to avoid causing offence while still trying to communicate what they desire. Examples include the use of indirect language, statements of general rules or questions. Recently, Ginsburg and colleagues³² have applied this theory to the written feedback provided by faculty to learners; they found the use of non-literal language was common in these circumstances and enabled faculty members to "save face". Politeness in interactions inevitably causes

confusion and can even be dangerously unhelpful in certain situations.³³ It was clear from our study that politely-expressed comments, or even silence, in the group did not adequately represent more firmly-held personal beliefs which became evident in the individual interviews.

Implications for Medical Education

It is understandable that most participants had strongly-held personal beliefs which kept them firmly rooted in the summative assessment culture. For both clinicians and students, it is likely that the prevailing assessment culture throughout school, university and their professional life had been based on high-stakes assessments, with little or no experience of an assessment for learning culture.

If institutions wish to change the assessment and feedback culture within their organisation, it will clearly be insufficient simply to attempt to implement assessment for learning approaches when, despite the evidence of the potential benefits, many stakeholders will be perplexed or resistant to change. Because of cultural conventions of politeness, faculty, clinicians and students may be reluctant to openly express their concerns or beliefs. Indeed, many of the beliefs may be so ingrained and taken for granted within the organisation that they may not be explicitly aware that their personal epistemologies would be an impediment for change.

In order to bring about a change in assessment culture towards one based on programmatic assessment or assessment for learning, the vital factor would appear to be a change in how both students and faculty conceptualise assessment. To accept change, they would need to stop believing in the primacy of summative high-stakes assessments. This requires a radical change in belief and the challenge of how to bring this about should not be underestimated.

When considering such a fundamental change in belief, perhaps we could learn from the field of science education. School science teachers are required to change children's firmly held preconceptions that the world is flat to an understanding that the earth is spherical. Simply presenting pupils with the apparently irrefutable evidence of a globe fails to convince students of the need to change their beliefs.³⁴ Instead, Vosniadou³⁵ argues for what she calls a synthetic models approach to conceptual change within science education. This approach argues that learners form pre-conceptions at an early stage, based on their experience of life. A number of pre-conceptions typically combine together to form a coherent if narrow explanatory framework, sometimes called naïve theory. Vosniadou³⁵ recognises that conceptual change is not a sudden replacement of one concept with another. Instead it is a slow process that involves a large network of inter-related concepts and which requires the development of new constructions that involve radical changes in personal beliefs. There is a risk with conceptual change that new conceptions are only partially understood, so that pre-conceptions are replaced by misconceptions. Vosniadou

argues for beliefs to be challenged so that they are not seen as undisputed scientific facts, but instead as hypotheses to be tested.

Although untested in medical education, Vosniadou's model appears a plausible proposal for bringing about the huge conceptual change necessary to support the introduction of programmatic assessment. Institutions enthusiastic to implement a change in assessment culture would need to understand the inter-related preconceptions maintaining belief in the summative assessment paradigm. Such preconceptions include the belief that exams are necessary to make students learn, the belief that summative assessments reassure the public that potentially unsafe doctors are prevented from practice and the belief that marking, numbers and grades are more important than feedback.

Although the difficulty of bringing about huge changes in conceptual thinking around assessment may make for somewhat depressing reading, it should be remembered that radical educational innovations can be successful. While the introduction of problem-based learning into medical undergraduate curricula has often been challenging, there are examples of very successful implementation.³⁶ In Québec, the enactment of change appeared to work as the faculty leaders built on the organisation's shared beliefs about the faults with the existing curriculum. This was linked to a gradual exposure (over several years) to educational changes occurring elsewhere, which meant that the institution's own proposed changes did not seem unduly radical.³⁶ Our study has demonstrated that stakeholders share common beliefs about problems with the current assessment culture, especially with regard to a lack of authenticity in clinical assessments, and there is therefore scope to explore these concerns further as one way of enhancing motivation to change.

There may also be lessons to learn from clinical settings, where there are daily struggles to encourage patients to change unhealthy lifestyles which are contributing to ill health. From the developing literature on health behaviour change, it is clear that listening carefully to patients, and taking time to explore their health beliefs in a non-judgemental manner, can help patients to resolve their own ambivalence about change.³⁷ Various health behaviour change theories have been used in a range of educational contexts to help researchers understand human behaviour in these settings, with promising results.³⁸ Our study has demonstrated the insights that can be gained when time is taken to carefully explore stakeholders' beliefs about assessment culture and their attitude to change.

Limitations

There are a number of limitations with our study. Studies using interviews and focus groups are inherently limited to considering only the participants' perspectives. It was conducted in a single institution and it is likely that participants from other sites would have expressed different beliefs. If the group had met on a number of occasions, interactions between stakeholders may well have changed over time. We do not claim to have achieved data saturation. However, the context in which the study took place is not unusual, as most

medical schools conduct summative assessments. Indeed, the medical school's record of innovation in the delivery of feedback may mean that our participants were more open to the possibility of change than might be the case elsewhere. Our findings also appear to resonate with the existing literature on the difficulty of bringing about organisational change.¹¹⁻¹³

Suggestions for Further Research

In this study we have explored personal and collective beliefs which would potentially hinder the implementation of a change in assessment culture. Further research is needed to investigate if these findings are replicated in other settings. More work is also required to understand how firmly-held beliefs about summative assessment may be challenged and modified.

Conclusion

This study has sought to understand the personal and collective beliefs which influence potential redesign of an assessment culture towards one which emphasises assessment for learning. We have shown that a variety of stakeholders hold common assumptions about the primacy of summative assessment. A lack of prior experience of alternative assessment cultures hinders the conceptualisation of radical change. In order to successfully implement a change in assessment culture, firmly-held intuitive beliefs about summative assessment will need to be challenged as a first step.

References

1. Shepard L. The role of assessment in a learning culture. *Educational Researcher* 2000; 29: 4-14.
2. Schuwirth LW, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 2011; 33: 478-485.
3. Van der Vleuten CPM, Schuwirth LW, Driessen E, Dijkstra J, Tigelaar D, Baartman LKJ, Van Tartwijk J. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012; 34: 205-214.
4. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007; 82: 493-502.
5. Dannefer E, Bierer B, Gladding SP. Evidence within a portfolio-based assessment program: What do medical students select to document their performance? *Medical Teacher* 2012; 34: 215-220.
6. Altahawi F, Sisk, B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: experience in a competency-based portfolio system. *Medical Teacher* 2012; 34: 221-225.

7. Bok HGJ, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LFH, Haarhuis, JCM, Van Beukelen P, Van der Vleuten CPM, Jaarsma DADC. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education* 2013; 13: 123.
8. Heeneman S, Oudkerk Pool A, Schuwirth LWT, Van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning – the theory versus practice. *Medical Education* 2015;49:487-98.
9. Dolmans D, de Grave W, Wolfhagen I, Van der Vleuten C. Problem-based learning: future challenges for educational practice and research. *Medical Education* 2005; 39: 732-741.
10. Dornan T, Scherpbier A, King, N, Boshuizen H. Clinical teachers and problem-based learning: a phenomenological study. *Medical Education* 2005; 39: 163-170.
11. Johnson G. Managing strategic change—strategy, culture and action. *Long range planning* 1992;25:28-36.
12. Scholes K, Johnson G. Exploring public sector strategy. Pearson Education; 2001.
13. Kavanagh MH, Ashkanasy NM. The Impact of Leadership and Change Management Strategy on Organizational Culture and Individual Acceptance of Change during a Merger. *British Journal of Management* 2006; 17, S81-S103.
14. McCoubrie P. Innovation in medical education: More than meets the eye. *The Clinical Teacher*. 2007 Mar 1;4(1):51-4.
15. Cook-Sather A. Unrolling roles in techno-pedagogy: toward new forms of collaboration in traditional college settings. *Innovative Higher Education* 2001; 26: 121-139.
16. Könings KD, Seidel T, Van Merriënboer JGG. Participatory design of learning environments: integrating perspectives of students, teachers and designers. *Instructional Science* 2014;42:1.
17. Könings KD, Brand-Gruwel S, Van Merriënboer JGG. Teachers’ perspectives on innovations: implications for educational design. *Teaching and Teacher Education* 2007; 23: 985-997.
18. Cheng L. Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education* 1999; 15: 253-271.
19. Könings KD, Brand-Gruwel S, Van Merriënboer JGG. Towards more powerful learning environments through combining the perspectives of designers, teachers and students. *British Journal of Educational Psychology* 2005; 75: 645–660.
20. Könings KD, Van Zundert MJ, Brand-Gruwel S, Van Merriënboer JJ. Participatory design in secondary education: Is it a good idea? Students’ and teachers’ opinions on its desirability and feasibility. *Educational Studies* 2007;33:445-65.
21. Cook-Sather A. Multiplying perspectives and improving practice: what can happen when undergraduate students collaborate with college faculty to explore teaching and learning. *Instructional Science* 2014;42:31-46.
22. Hofer BK. Personal epistemology research: implications for learning and teaching. *Journal of Educational Psychology Review* 2001; 13: 353-383.
23. Coxhead L, Rhodes T. Accounting for risk and responsibility associated with smoking among mothers of children with respiratory illness. *Sociology of Health and Illness* 2006; 28: 98-121.
24. Sweeney K. Complexity in Primary Care: understanding its value. Oxford: Radcliffe Publishing 2006.
25. Harrison CJ, Könings KD, Schuwirth L, Wass V, Van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education* 2015;20:229-45.

26. Harrison CJ, Könings KD, Dannefer EF, Schuwirth LW, Wass V, Van der Vleuten CP. Factors influencing students' receptivity to formative feedback emerging from different assessment cultures. *Perspectives on Medical Education* 2016;5:276-84.
27. Harrison C, Könings K, Molyneux A, Schuwirth L, Wass V, Van der Vleuten C. Web-based feedback after summative assessment: how do students engage? *Medical Education* 2013; 47: 734-744.
28. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Medical Teacher* 2017;39:44-52.
29. Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Medical Teacher* 2015;37:323-6.
30. Janss R, Rispens S, Segers M, Jehn KA. What is happening under the surface? Power, conflict and the performance of medical teams. *Medical Education* 2012;46:838-49.
31. Brown P, Levinson SC. *Politeness: Some universals in language usage*. Cambridge University Press; 1987.
32. Ginsburg S, Van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Advances in Health Sciences Education* 2016;21:175-88.
33. Bonnefon JF, Feeney A, De Neys W. The risk of polite misunderstandings. *Current Directions in Psychological Science* 2011;20:321-4.
34. Vosniadou S, Skopeliti I, Ikospentaki K. Reconsidering the role of artifacts in reasoning: Children's understanding of the globe as a model of the earth. *Learning and Instruction* 2005;15:333-51.
35. Vosniadou S. Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. In: *Second international handbook of science education*. Springer Netherlands; 2012. p 119-130.
36. Des Marchais JE, Bureau MA, Dumais B, Pigeon G. From traditional to problem-based learning: a case report of complete curriculum reform. *Medical Education* 1992;26:190-9.
37. Rollnick S, Miller WR, Butler CC. *Motivational interviewing in health care: helping patients change behaviour*. New York: The Guildford Press; 2008.
38. Cilliers F, Schuwirth L, Van der Vleuten C. 13 Health behaviour theories: a conceptual lens to explore behaviour change. *Researching Medical Education* 2015;8:141.

Chapter 6

Discussion

This thesis has sought to explore the challenges of receiving feedback in the context of high-stakes assessments. In particular, there were several overarching research questions which this thesis has addressed:

1. How do students engage with feedback in a high stakes context?
2. What are the influences and barriers to feedback receptivity, both at an individual and institutional level, in this context?
3. Can we change the assessment culture to one which prioritises and supports learning from feedback? What is stopping us?

This chapter will start by considering the answers that were found to these questions, linking the findings to other work and discussing the implications for medical education. The strengths and limitations of the thesis will be considered and suggestions for further research in this area will be proposed.

How do students engage with feedback in a high stakes context?

There has been little previous work on student engagement with feedback following summative assessment. In one study which considered feedback following written assessment, fewer than half the students took up an opportunity to view their feedback.¹ In our study, reported in chapter 2, we found that almost all students visited the website which we developed to provide feedback after summative OSCE examination. However, although results from our study seem better at first glance, they are not reassuring; most students appeared to only superficially engage with the feedback as they typically spent only a very short time on each webpage. Use of the feedback website was also not homogenous, with a wide range in the number of webpages viewed and in the number of times the website was visited. The students who had only just passed the OSCE were the least engaged. This was a novel finding, not previously described. The data obtained could not explain why this might be and the lack of similar findings in the literature meant there was a need to explore why students engaged in the way they did (or did not) with the feedback website. The study also sought to explore links between students' learning-related characteristics, such as goal orientation and motivation, and their use of feedback. Overall, no strong links were seen. This is perhaps surprising given the well-established links between learning-related characteristics and feedback when it is given in a context other than summative assessment.^{2,3} It was not possible to establish from this quantitative study whether this was because the context was a summative rather than formative assessment, or because other factors were more important. The overall conclusion from this study was that it is not sufficient simply to provide feedback and expect students to make full use of the

opportunities provided, but it was unable to provide a clear understanding of the factors influencing feedback use.

What are the influences and barriers to feedback receptivity, both at an individual and institutional level, in the context of high-stakes assessments?

As the first study was unable to explain why learners may, or may not, engage with feedback in the context of summative assessment, the study in Chapter 3 therefore employed in-depth interviews to explore the individual barriers to feedback receptivity. One of the main findings was the perceived power of the summative assessment culture. The fear of failure and the subsequent “punishment” of needing resits loomed large in students’ minds. This meant that exams were seen as a hurdle, not a stepping stone to future learning. Students who had passed the assessment felt that they had achieved their goal which meant that they did not perceive a “feedback gap” between their own performance and the required standard. As a result, they were not motivated to meaningfully engage with the feedback. This links with well-established literature that has demonstrated that motivation is vital to enable an appropriate response to feedback.²

Students in the study felt a disconnection between the assessments and clinical workplace learning; this was reinforced by the attitudes of clinical teachers and family members. This finding resonated with situativity theory, which proposes that educators need to consider carefully the context in which information will (or will not) be used, rather than just concentrating on the content of the information to be imparted.⁴ The theory proposes that authentic learning environments should help the learning process.⁴ Unfortunately, the high-stakes punitive assessment hurdles are perceived by student as existing in a separate universe to the authentic learning which takes place within the clinical workplace. The implication of these findings is that simply concentrating on trying to improve feedback delivery, by exhorting faculty to provide more, or better-quality, feedback, is doomed to fail as this approach fails to recognise the context in which the assessment and feedback take place. Similarly, urging students to reflect more carefully on the feedback is not going to be sufficient on its own, unless factors such as learners’ motivation and the disconnection from authentic learning are addressed. Although the study in chapter 3 set out to explore the individual barriers to feedback receptivity, it therefore demonstrated that factors within a medical school’s assessment culture were very significant. As a result, the focus should be on the development of a feedback culture, as others have proposed.⁵

The study in Chapter 4 therefore sought to understand the factors within medical school assessment systems which influenced feedback receptivity in more detail. In a related finding to the previous study, assessments perceived as inauthentic hindered receptivity to

feedback. By contrast, students positively valued assessments and feedback if they were clearly relevant and aligned with their future work. Although grades provided a level of superficial reassurance that students were “on track” to graduate as a doctor, failure to provide grades appeared ultimately more helpful to students. Although left feeling initially uncertain, the lack of grade labels in time enabled students to appreciate the finer nuances of individual strengths and weakness, both in them and among their peers. This links to recent work which demonstrated the complex relationship students have with grades in a workplace setting.⁶ In that study, the majority of students wanted to receive grades on their assessments as they provided calibration about their progress towards graduation. However, in some students, satisfactory grades appeared to induce a degree of complacency, and nearly a quarter of students expressed a preference for feedback without grades.⁶

An interesting finding from the study in Chapter 4 was that openness to feedback was promoted by offering choice and independence within assessment systems. When assessment systems provided elements of choice, enabling students to demonstrate their knowledge to themselves or their tutors, students’ agency was enhanced. Within feedback, agency was enhanced by enabling students to act independently in accessing and using feedback. These findings had not previously been clearly demonstrated in the literature. Although active involvement by students in the learning process has increasingly been accepted, it has proved much more difficult to achieve this transformation in assessment and feedback processes, despite enthusiastic advocates.⁷ It could be argued that there are significant risks to patients if students were given excessive choice in the assessment system and were free to ignore uncomfortable feedback. As with the finding from the first study, it was clearly not sufficient to simply provide feedback and expect students to make appropriate use of it; there was a need for mentoring. Mentoring helped students to interpret feedback and could support the fostering of agency if the scaffolding was reduced as time went on through the course. The study reinforces the potential learning benefits of mentoring in a supportive relationship, linked with authentic assessment and feedback, which have been demonstrated in studies of longitudinal integrated clerkships.⁸ At times, mentors adopted a more paternalistic approach, which was not helpful for promoting agency. This resonates with the literature on the challenges facing problem-based learning tutors.⁹

The studies in Chapters 3 and 4 demonstrated how assessment cultures have remained rooted in a behaviourist approach to medical education: students are rewarded for passing and punished for failing assessments. The studies also supported a move towards a more constructivist approach to assessment, empowering the students to take a more active role. Other constructivist elements were a clear call for more authenticity in assessments and the benefits of mentoring, with gradual descaffolding as the course progresses. A behaviourist approach to learning in medical education, relying on didactically-delivered lectures to passive students, now seems increasingly outdated, but a behaviourist approach to assessment has mostly persisted. Even in institutions which have adopted active approaches

to learning, the associated assessments have often rewarded superficial learning strategies, which students understandably adopt.^{10, 11} Just as the concept of a patient as a passive recipient of healthcare seems obsolete in the internet era, so the idea of a student passively accepting the rewards and punishments of an assessment system based on behaviouristic principles appears archaic and increasingly difficult to defend. This may at least partly explain why student surveys almost always demonstrate much less satisfaction with assessment and feedback than with other aspects of the course.¹²

These studies also add to the growing literature demonstrating the adverse effects of summative assessment on learning.¹³⁻¹⁵ While the powerful and often harmful effects that high-stakes assessments can have on pre-assessment learning in medical education have already been established,¹³⁻¹⁵ the studies demonstrate how summative assessment can stifle the post-assessment learning effects. The findings also support Watling's contention that the culture within medical education appears to value the attainment of a state of competence in order not to need further observations and feedback, rather than a continued striving for further improvement which is the norm in some other learning cultures.¹⁶⁻¹⁸ It is apparent that emotions can often act as a barrier to feedback receptivity, confirming earlier work.^{19, 20}

Can we change the assessment culture to one which prioritises and supports learning from feedback? What is stopping us?

The previous studies highlighted the need to change the assessment culture, from one which focusses on assessment of learning to one which supports assessment for learning, in order to enhance feedback receptivity. The study described in Chapter 5 used participatory design methods to explore how a mixed group of stakeholders might attempt to change the assessment culture to one in which students made more use of post-assessment feedback. Participants shared common assumptions and beliefs about the importance of the summative assessment paradigm. Their beliefs seemed heavily influenced by their own prior experiences of assessment, were deeply-rooted and largely not changed by arguments from others in the group discussion. As a result, changes that were proposed were typically small and still congruent with the summative assessment paradigm. Strongly-held beliefs were often in conflict; for example, a tension existed between the need for greater authenticity in assessments and the need for standardisation and reliability. Van der Vleuten's work on the need for compromise in assessment clearly remains very relevant.²¹

This study demonstrated the enormity of the challenge for assessment programme designers who wish to bring about a change in assessment culture and helps to explain why previous attempts to introduce assessment innovations have not always been straightforward.²² The study's findings should not, however, be so surprising. The dominance of the prevailing

paradigm in preventing radical change is very consistent with the literature on organisational change.²³ The faculty development literature has clearly demonstrated that teacher behaviour is strongly influenced by their underlying beliefs and prior experiences.²⁴ Short-term courses focussing on developing a recipe of teaching skills have very limited impact, but sustained, long-term interventions have shown promising results.^{25, 26} Such interventions can include exposure to alternative beliefs and conceptions about teaching, through discussion with peers and facilitators, followed by a facilitated critical self-analysis of participants' existing beliefs.²⁵ Although participatory design approaches can be very helpful, the intervention in our study was very short-term and therefore could not be expected to make much impact. However, longer-term interventions may be helpful. In our study, participants' prior experiences seemed particularly influential, so exposure to alternative assessment models, perhaps in centres which have successfully implemented programmatic approaches to assessments, may prove helpful.

Implications for Medical Education

The answers to the research questions are therefore quite challenging: although the practical difficulties of delivering feedback in a high-stakes assessment context can be overcome, this does not mean that learners will be receptive to the feedback available to them. Teasing out the logical inconsistencies brings to mind the so-called “catch-22” logic, a term coined by the author Joseph Heller in his book of the same name²⁷ Summative assessment is necessary to assure regulators and the public that learners have reached a minimum standard of competence. This minimal competence is not sufficient, hence the need for feedback to help plan further learning. Although feedback can be a very powerful tool for learning, students appear largely unreceptive to feedback when it is given in the context of high-stakes assessments. Logically this should lead to a change away from the summative assessment culture in order that feedback can be valued. However, when we consider changing the assessment culture, stakeholders are very reluctant as summative assessment is seen as so important to assure regulators and the public of a minimum standard of competence. And so the circular argument begins again.

Although these findings may cause discomfort, they should not be used as an excuse to maintain the status quo. Single-shot high-stakes assessments remain inherently flawed.²¹ and the results deliver a message that is at best misleading.²⁸ Although many students appeared unreceptive to feedback in this context, it does not invalidate the strong evidence of the powerful effects of feedback, when the conditions are right.^{2, 3} One of the problems of feedback delivered in a high-stakes context is that there is a lack of a perceived gap between learners' observed performance and the required standard; if students have passed a summative assessment they see this as, by definition, having reached the required minimum standard. Whatever the challenges, the assessment culture still needs to be

modified so that the undoubted benefits of feedback can be harnessed. Of course, each medical school's assessment culture is unique, so the extent to which high-stakes assessments are perceived as punitive and dominant will vary from school to school.

When one appears unclear on finding the way forward, it is often worth looking back; history can offer us interesting insights. One of the key concerns for stakeholders is that new programmes of assessment seem unfamiliar and therefore somewhat radical and fear-provoking. But are the proposals really that new and radical? More than one hundred years ago, Sir William Osler clearly understood the adverse effects arising from the dominant influence of a summative assessment culture on medical education.²⁹ His eloquent words deserve detailed scrutiny. In an address to a London medical school, he said:

We make the examination the end of education, not an accessory in its acquisition. The student is given the impression that he is in the school to pass certain examinations and I am afraid the society in which he moves grinds this impression into the soul.

Among his suggested remedies are adding elements of what we would now call an assessment for learning culture, while not completely abandoning assessment of learning. His proposals include frequent observations and judgements made in the workplace. These assessments would be low stakes in nature, but count towards the final examination mark. This would result in the final, formal examination being an “amplification” of the multiple low stakes assessments. He even advocated substituting qualitative assessments for quantitative estimates. Although he does not explicitly use the term feedback, he cannot be criticised for this, as the term was not in use until much later. With regard to students persistently struggling, he advocated honest dialogue, recognising that it was kinder for them to change career at an early stage rather than “struggle on painfully and submit to humiliation”. He had even thought about the need to foster students' agency, quoting a former colleague approvingly:

Let us emancipate the student, and give him (sic) the time and opportunity for the cultivation of his mind, so that in his pupillage he shall not be a puppet in the hands of others, but rather a self-relying and reflecting being.

His proposed assessment system would, he foresaw, have great benefits on students' learning:

[This] would enable him (sic) to seek knowledge for itself, without a thought of the end, tested and taught day by day, the pupil and teacher working together on the same lines.

In recognition of successful change management strategies, he made clear that examiners would also benefit from the proposed changes:

It is much nicer to watch the gradual growth of a student's knowledge and to get it out retail day by day than to drag it out wholesale at set times.

While accepting there were potential implementation challenges, he was nevertheless optimistic:

Reform is in our hands and should not be far off.

Sadly, the results of this thesis demonstrate that the need for reform is still largely awaited, rather than a footnote in history. Possible ways forward will now be considered, using Osler's wisdom and historical perspective to guide us where necessary.

One obvious solution would be to make the summative assessments more challenging, by redefining medicine's culturally accepted notions of acceptable minimum competence for graduation to something more stringent. In this way, when students perceive no feedback gap between their performance and the required standard, they have at least already achieved a higher initial level of competence. However, this proposal is flawed on a number of levels. Firstly, it would be likely to lead to a higher initial number of students failing high-stakes assessments at the first attempt. This was one of the main problems with the assessment system in Osler's time which led him to propose his changes; in his time failure rates of over 40% in the main high-stakes assessments were commonplace.²⁹ Osler was clear on the distress this caused to students and was unsparing in his criticism of examiners who appeared to enjoy humiliating candidates. On a practical level and in the current economic climate, institutions would find it hard to support increased numbers of students through a remediation process. Irrespective of the logistical challenges, it is clear that medicine has much to do to improve the process of remediation.³⁰ More fundamentally, making high-stakes decisions on single assessments is inherently flawed; some candidates will fail who should have passed, and vice versa.³¹ The process of defining minimal competence is fraught with difficulty, with a lack of clear agreement on hypothetical students who just about deserve to pass.³² As usual, Osler²⁹ was here before us, acknowledging the difficulties in defining essential minimum requirements. Perhaps the most important flaw with this proposed solution is that it resides within the existing punitive behaviourist paradigm of existing high-stakes assessments; students are rewarded for passing and punished for failing. Raising the bar on an individual assessment would simply make the stakes even higher, so the emotional barriers to reception of feedback would be likely to be amplified. Instead, solutions which are situated within a constructivist approach to learning and assessment would seem more appropriate.

If raising the bar of an individual summative assessment does not provide an easy answer, another logical step would be to combine several assessments together making progress decisions on the cumulative performance in all the assessments. One way of achieving this is through the Progress Test. This is an assessment which seeks to evaluate students' knowledge of the learning objectives covering the whole curriculum.³³ Administered

several times per year, it is designed to encourage deeper learning strategies and discourage last-minute superficial learning. The progress test is just one approach; other cumulative approaches to assessment have been proposed as a way of encouraging retention of previously learnt material, with some promising preliminary results.^{34, 35} Unfortunately, it is not inevitable that any of these approaches will have the desired positive learning effects. The overall design of the assessment system can prevent these benefits being realised. The frequency of administration of the progress test, and the use of other end-of-block assessments can influence whether students perceive the test as an aid to their learning or is instead regarded as just another hurdle.³⁶ A recent study demonstrated that students made less use of feedback available after progress tests if summative decisions were made on the cumulative performance in the last four tests. By contrast, when the progress test was embedded in a more comprehensive programme of assessment, and the results of the tests not used summatively but were instead part of a broader portfolio system, supported by long-term mentoring, students made significantly more use of feedback.³⁷ The important implication of these studies is that it is not the assessment instrument (such as the progress test) which inevitably fosters an assessment for learning approach; instead the critical issue is how the instrument is used within the overall assessment culture.³⁸

It therefore appears that the way forward is by careful design of the whole assessment programme, paying careful attention to its educational impact as well as the more traditional considerations of validity and reliability. The field of medical education is not alone in struggling to reach an effective integration of these concepts; the wider educational establishment is also trying to determine how assessment programme design can foster learning.^{39, 40} In order to respond to assessment-related feedback, there is a need to align learning theories with assessment paradigms.⁴¹ For many years the dominant paradigm has been psychometric, with its reliance on the use of quantitative data, manipulated in standardised and 'objective' ways, in order to justify summative decisions. The knowledge that is assessed is generally regarded as independent of the context. A more promising assessment paradigm arises from the sociocultural perspective.^{7, 42} In this paradigm, learning is assessed in a particular context. The relationship between learners and the wider community is also taken into account. Students are more actively involved in the assessment process.^{7, 42} Although this assessment paradigm has only become more prominent in recent years, aspects of it are not entirely new; Osler clearly understood the importance of assessment in context.²⁹

When designing a programme of assessment, medicine arguably faces particular challenges because of the need for regulators to reassure the public that their doctors are competent. Doctors are usually comfortable with the psychometric method as the predominant paradigm, which has been valued in medical research, typically favours a quantitative and positivistic approach. Although the importance of context specificity in medical assessment has been recognised for many years, the sociocultural assessment paradigm remains unfamiliar to most doctors. However, the practice of medicine is frequently recognised to

be an art as much as a science.⁴³ This is perhaps most apparent in the specialty of general practice, where the sociocultural context is seen as crucial. In the previous chapter, the strength of intuitive beliefs about assessment paradigms was demonstrated and some suggestions as to how these may be challenged were considered. Another potential way forward is by making the proposed new paradigm feel less unfamiliar. One method of achieving this is by drawing analogies to a context with which clinician-educators are very familiar with, namely healthcare. The lessons that can be learned from healthcare, especially from the field of general practice or family medicine, will now be explored.

The importance of patients' beliefs in the daily 'bread and butter' of general practice cannot be overstated. Often the beliefs may be naïve and misguided, but they are firmly held. They come from patients' prior experience of healthcare, for themselves or their family members. If explored carefully, a belief that seems implausible often has very rational and understandable origins; for example, a family member previously presented with atypical symptoms, which meant that a rare illness was not recognised. A skilful general practitioner (GP) doesn't just ignore the beliefs, as that is neither productive nor efficient and neglects the all-important doctor-patient relationship.⁴⁴ Instead, a good GP seeks to negotiate a treatment that is both compatible with the patient's beliefs and attuned with the scientific and medical evidence of what is needed in this case. Normally a compromise can be reached. At the same time, some beliefs need to be gently challenged if they are not compatible with a safe clinical outcome. Are there lessons to be learnt for medical education? Beliefs in the benefits of summative assessment may seem naïve to those fully conversant with the developing literature, but they are firmly held and, if carefully explored, it is clear that they too have a rational basis. There is a desire for rigour and upholding of standards to enter the medical profession and there is a perception that newer forms of workplace assessment lack objectivity. Skilful designers of assessment programmes should therefore try to acknowledge and accommodate those beliefs where possible. When the beliefs are not compatible with ensuring fairness for students and ultimately a good clinical outcome for patients, then a dialogue is needed to challenge these beliefs, within an attitude of mutual respect.

As others have pointed out, there are other potential similarities between the promotion of formative assessment and good clinical care, especially in general practice.⁴⁵ Mentoring seems very beneficial in helping learners to interpret feedback but appears hard to implement in practice; this is very similar to continuity of care which is much valued by patients although practices find it hard to deliver with limited resources.⁴⁶ However, a key benefit is that GPs also value continuity of care as it makes the job more meaningful and satisfying; educators may well find long-term mentoring similarly rewarding.⁴⁷ There are also arguments that it is more efficient as resources can be saved by avoiding unnecessary or redundant investigations and referrals.⁴⁸ Mentoring in medical education might also ultimately save resources by intervening earlier in a learner's career with targeted remediation, rather than waiting for failure in end-of-year assessments followed by costly

resits; the old English proverb, ‘a stitch in time saves nine’ has long highlighted the benefits of avoiding procrastination. If the students become ‘trained’ at being receptive to feedback, interpreting it and then changing their behaviour, it is at least plausible that this will help them in their career to avoid expensive and dangerous clinical mistakes. Just as compromises in achieving continuity in clinical practice have been proposed, there could be other ways of achieving some of the aims of mentoring, perhaps by using a small team of mentors communicating closely with each other.⁴⁶

Emotions can be a potent barrier to patients being open to receive information in consultations, especially those that involve the reception of bad news.⁴⁴ However, once the intensity of emotions has settled, patients are often extremely interested in receiving more information to help them understand their condition and plans for more treatment. So emotions in themselves should not be an insuperable barrier to prevent students being receptive to feedback information. However, the crucial difference in education is that the ‘diagnosis’ in summative assessment terms is pass or fail. If the diagnosis is ‘pass’, then there is no incentive for students to need more ‘treatment’. If the diagnosis is ‘fail’, then they do have an incentive to engage, and there is some evidence from the first study in this thesis that many failing students will engage with feedback. We therefore need to find a way in which the diagnosis is not simply a pass but a more nuanced description of strengths and weaknesses so that students have an incentive to seek more information to help them plan their ‘treatment’.

It may therefore be possible for assessment programme designers to employ some of these insights from the clinical perspective while trying to change the assessment culture to one that is more supportive to students receiving feedback in a meaningful way. Time should be taken to explore beliefs in a way that fosters mutual respect, in order to try to reach compromises which are reasonably acceptable to all stakeholders. While accepting that the challenges of implementing long-term mentoring are very significant, there is great potential that clinicians and students would subsequently find the assessment system more rewarding in the end, with beneficial effects on intrinsic motivation.⁴⁷

Strengths and Limitations

This thesis has a number of strengths. Firstly, it seeks to deal with an important and relevant problem: no learner can avoid assessment, and high-stakes assessments remain the norm at both undergraduate and postgraduate levels. It is also original, as there has been very little empirical work done to date on feedback receptivity in the context of summative assessments. The research was also rigorous, employing a variety of different methodologies, which has aided triangulation of emerging concepts. The involvement of

three different institutions, in three separate countries, with very diverse assessment cultures helps to enhance the robustness of the findings.

There are equally limitations, many of which have already been detailed in the reports of the individual studies. Although other countries were involved for one study, the other studies took place in a single institution and, in two cases, with a single year group. This may limit the degree to which the findings can be generalised. Further work is needed to replicate the findings. It is unclear whether the outcomes of the research will be stable over time; more long-term research is needed. The main aim of the thesis, however, was to gain a clearer understanding of the issues around combining formative feedback with summative decision-making in an assessment system, especially from the perspective of the students, so the generalisability and stability of the findings over time is not necessarily the main concern.

Implications for further research

The research reported in this thesis has helped to clarify the challenges of receiving feedback within the context of summative assessment and why it is likely to be challenging to change the assessment culture towards one which is more supportive of engagement with feedback. It is increasingly apparent that a paradigmatic shift is needed in the assessment and feedback culture. Future work should include explorations of approaches to better understand, then modify or challenge, stakeholders' beliefs about assessment, as this appears to be an important precursor to change. Although several articles have been published in recent years arguing for a coherent, integrated programme of assessment, with a focus on formative feedback,^{45, 49, 50} there is a need to research more closely the actual implementation of change in the assessment culture. There are now a number of examples where a programmatic approach to assessment appears to have been implemented with a significant degree of success.⁵¹⁻⁵³ Further qualitative research in these contexts could provide useful insights into the factors which helped implementation. Equally, and perhaps even more importantly, insights could be gained from institutions which have tried to introduce an assessment for learning culture but in which the implementation has been less successful. Models of health behaviour change could provide a useful theoretical lens through which to look at this issue.⁵⁴

Successful implementation will also depend on sufficient resources. Assessments in medical education, especially practical skills assessments such as OSCEs, are extremely costly and there is a need to ensure assessment programmes provide appropriate value for money.⁵⁵ There has so far been very limited work on the costs of assessment programmes. An argument can be made that programmatic assessment can be affordable, provided that tough choices are made.⁵⁶ For example, in the later years of a medical school programme,

OSCEs may need to be dropped in favour of more workplace-based assessments. This has the added benefit of helping to ensure greater authenticity in the assessments. Empirical work on the cost-effectiveness of programmes could usefully learn from similar work in the healthcare field.⁵⁷

The need for sufficient resources is not a purely financial argument; many schools may struggle to muster sufficient human resources to provide the high-quality mentoring which appears so important. As a consequence, compromise methods to implement important aspects of enhancing feedback receptivity should be considered; for example, might students engage satisfactorily with online mentoring as a way of helping them to engage with feedback? The long-term benefits, and possible harms, of mentoring in this context should be considered. Does mentoring students about their response to feedback lead to improved performance in assessment, or long-term performance as a doctor?

Conclusion

This thesis sought to challenge an old assumption in medical education, namely that a dichotomy should be maintained between summative and formative assessment. The challenging conclusion of the thesis is that, in many cases, learners will not be receptive to feedback that may be available after high-stakes assessments. The summative assessment culture is dominant and acts to stifle the learning potential available from assessments. Its dominance also makes any change of culture very challenging to implement, however necessary the change may appear. Possible ways of enhancing post-assessment feedback receptivity have been proposed. Medical education needs a paradigm shift in the assessment and feedback culture. This should benefit teachers, students but, crucially, also patients who have much to gain from doctors who are open to receiving, interpreting and acting on feedback. For the final words, Osler²⁹ yet again offers wisdom:

The truth is, we have outrun an educational system framed in simpler days and for simpler conditions. The pressure comes hard enough upon the teacher, but far harder upon the taught, who suffer in a hundred different ways. To help you realise this pressure and to suggest measures of relief are the objects of this address.

They have also been the objects of this thesis.

References

1. Sinclair HK, Cleland JA. Undergraduate medical students: who seeks formative feedback?. *Medical Education* 2007;41:580-2.
2. Hattie J, Timperley H. The power of feedback. *Review of Educational Research* 2007;77:81-112.
3. Shute VJ. Focus on formative feedback. *Review of Educational Research* 2008;78:153-89.
4. Durning SJ, Artino AR. Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. *Medical Teacher* 2011;33:188-99.
5. Archer JC. State of the science in health professional education: effective feedback. *Medical Education* 2010;44:101-8.
6. 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. *Medical Education* 2015;49:307-20.
7. Rust C, O'Donovan B, Price M. A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education* 2005;30:231-40.
8. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Medical Education* 2013;47:362-74.
9. Dornan T, Scherpbier A, King N, Boshuizen H. Clinical teachers and problem-based learning: a phenomenological study. *Medical Education* 2005;39:163-70.
10. Savin-Baden M. Understanding the impact of assessment on students in problem-based learning. *Innovations in Education and Teaching International* 2004;41:221-33.
11. Struyven K, Dochy F, Janssens S. Students' perceptions about new modes of assessment in higher education: A review. In: Segers M, Dochy F, Cascalier E (Eds). *Optimising new modes of assessment: In search of qualities and standards*. Springer Netherlands; 2003. p 171-223.
12. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education* 2013;38:698-712.
13. Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, Van der Vleuten CP. The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education* 2010;15:695-715.
14. Cilliers FJ, Schuwirth LW, Herman N, Adendorff HJ, Van der Vleuten CP. A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education* 2012;17:39-53.
15. Cilliers FJ, Schuwirth LW, Van der Vleuten CP. A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Medical Education* 2012;12:1.
16. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Medical Education* 2013;47:842-50.
17. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education* 2013;47:585-94.
18. Watling C, Driessen E, Van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Medical Education* 2014;48:713-23.
19. Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education* 2012;17:15-26.

20. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education* 2008;13:275-88.
21. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education* 1996;1:41-67.
22. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, Van Beukelen P, Van der Vleuten CP, Jaarsma DA. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education* 2013;13:1.
23. Johnson G. Managing strategic change—strategy, culture and action. *Long Range Planning* 1992;25:28-36.
24. Gibbs G, Coffey M. The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education* 2004;5:87-100.
25. Ho A, Watkins D, Kelly M. The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education* 2001;42:143-69.
26. Knight AM, Carrese JA, Wright SM. Qualitative assessment of the long-term impact of a faculty development programme in teaching skills. *Medical Education* 2007;41:592-600.
27. Heller J. *Catch 22*. New York City: Simon and Schuster; 1961.
28. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, Regehr G. Towards a program of assessment for health professionals: from training into practice. *Advances in Health Sciences Education* 2016;21:897-913.
29. Osler W. Examinations, examiners and examinees. *The Lancet* 1913;182:1047-1050.
30. Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: theoretical and methodological insights from a systematic review. *Medical Education*.2013;47:242-51.
31. Van der Vleuten CP. Revisiting ‘Assessing professional competence: from methods to programmes’. *Medical Education* 2016;50:885-8.
32. Boursicot K. Setting standards in a professional higher education course: defining the concept of the minimally competent student in performance based assessment at the level of graduation from medical school. *Higher Education Quarterly* 2006;60:74-90.
33. Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Medical Teacher* 2010;32:451-5.
34. Swanson DB, Holtzman KZ, Butler A. Cumulative achievement testing: Progress testing in reverse. *Medical Teacher* 2010;32:516-20.
35. Kerdijk W, Tio RA, Mulder BF, Cohen-Schotanus J. Cumulative assessment: strategic choices to influence students’ study effort. *BMC Medical Education* 2013; 13: 1.
36. Wade L, Harrison C, Hollands J, Mattick K, Ricketts C, Wass V. Student perceptions of the progress test in two settings and the implications for test deployment. *Advances in Health Sciences Education* 2012;17:573-83.
37. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Medical Teacher* 2017;39 :44-52.
38. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Medical Education* 2016;50:704-6.

39. Knight PT. The value of a programme-wide approach to assessment. *Assessment & Evaluation in Higher Education* 2000;25:237-51.
40. Gibbs G, Dunbar-Goddet H. Characterising programme-level assessment environments that support learning. *Assessment & Evaluation in Higher Education*. 2009;34:481-9.
41. Van der Kleij FM, Vermeulen JA, Schildkamp K, Eggen TJ. Integrating data-based decision making. *Assessment for Learning and diagnostic testing in formative assessment. Assessment in Education: Principles, Policy & Practice* 2015;22:324-43.
42. Moss PA, Pullin D, Gee JP, Haertel EH. The idea of testing: Psychometric and sociocultural perspectives. *Measurement: Interdisciplinary Research and Perspectives* 2005;3:63-83.
43. Dixon M, Sweeney K. *The Human Effect in Medicine: Theory, Research and Practice*. Abingdon: Radcliffe Medical Press; 2000.
44. Silverman J, Kurtz S, Draper J. *Skills for communicating with patients*. CRC Press; 2016.
45. Konopasek L, Norcini J, Krupat E. Focusing on the Formative: Building an assessment system aimed at student growth and development. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2016; 91:1492-7.
46. Freeman G, Hughes J. *Continuity of care and the patient experience*. London: The King's Fund. 2010.
47. Driessen E, Overeem K. Mentoring. In: Walsh K. *Oxford Textbook of Medical Education*. Oxford: Oxford University Press; 2013. p265-274.
48. Hjortdahl P, Borchgrevink CF. Continuity of care: influence of general practitioners' knowledge about their patients on use of resources in consultations. *BMJ* 1991;303:1181-4.
49. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 2011;33:478-85.
50. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, Van Tartwijk J. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34:205-14.
51. Driessen EW, Van Tartwijk J, Govaerts M, Teunissen P, Van der Vleuten CP. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Medical Teacher* 2012;34:226-31.
52. Dannefer EF. Beyond assessment of learning toward assessment for learning: educating tomorrow's physicians. *Medical Teacher* 2013;35:560-3.
53. Li SA, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) Through the years: residents' experience of an evolving feedback culture created by a programmatic, workplace-based assessment system over a three-year period. *Academic Medicine* 2016;91:S10-11.
54. Cilliers F, Schuwirth L, Van der Vleuten C. Health behaviour theories: a conceptual lens to explore behaviour change. In: Cleland J, Durning SJ (Eds). *Researching Medical Education*. Chichester John Wiley & Sons Ltd; 2015. p141-153.
55. Brown C, Cleland J, Walsh K. The costs of medical education assessment. *Medical Teacher* 2016;38:111.
56. Van der Vleuten CP, Heeneman S. On the issue of costs in programmatic assessment. *Perspectives on Medical Education* 2016;5:303-7.
57. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. *JAMA*. 1996;276:1339-41.

Summary

**Feedback in the context
of high-stakes assessment:
Can summative be formative?**

Summary

Feedback can be an extremely powerful phenomenon, one of the most powerful ways to stimulate performance improvement. Unfortunately, it frequently fails to reach its potential. Reasons for this include poor delivery of feedback, problems with the reception of feedback, or because learners are unable to translate the feedback into actual learning activities that would lead to positive change. The role of feedback within the context of summative assessment has previously received little attention. This thesis therefore set out to consider the challenges of receiving feedback within this context, and in particular to challenge the common assumption that assessments have either a formative or a summative purpose and that it is difficult for a single assessment to serve both purposes at the same time.

There are several reasons why it would seem sensible to combine the formative and summative purposes of assessment. High-stakes assessments typically create large amounts of data regarding a student's strengths and weaknesses, but this information is reduced down to a binary pass-fail decision on progression to the next stage. While this is an understandable approach, it has the potential to mislead as much rich information about a learner has been lost in the process, and fails to prepare students for independent self-regulated learning. Detailed information from assessments has the potential to support students to identify their own strengths and weaknesses. Practical assessments in medical education are extremely expensive, so it also makes financial sense to maximise the value which can be obtained from them.

On the other hand the delivery of feedback in these circumstances presents practical challenges, with a need for significant faculty resources, although these can be potentially overcome, in part by the use of information technology. However, learners' receptivity to the feedback is more challenging; it is uncertain whether they will engage with the feedback in a meaningful way in this context. There are potential barriers at both the individual and cultural level to the reception of feedback, but these have not been studied in the context of summative assessment. Summative assessment is known to have adverse effects on other aspects of the learning process, so it is plausible that it could have undesirable effects on feedback receptivity. Alternatives to the summative assessment culture towards one based on an assessment for learning have been proposed, but there is a lack of empirical evidence as to how this can be successfully implemented in a way that engages relevant stakeholders.

This thesis therefore sought to address three principal research questions:

1. How do students engage with feedback in a high stakes context?
2. What are the influences and barriers to feedback receptivity, both at an individual and institutional level, in this context?

3. Can we change the assessment culture to one which prioritises and supports learning from feedback? What is stopping us?

The study described in **Chapter 2** sought to address the first main research question by using quantitative methodology to evaluate the extent to which medical students engaged with feedback after a high stakes clinical assessment. To address the problems of limited faculty resources and the need for timely feedback delivery, feedback information was provided in a web-based format. As well as helping with feedback delivery, the web-based format also enabled students' use of the information to be monitored. This website usage data was linked to students' performance in the assessment and their responses to a questionnaire concerning learning-related student characteristics. As an additional feature, the website provided guidance to learners on formulating an action plan following study of the feedback. Our analysis demonstrated that most students engaged with web-based feedback after a summative OSCE, but at a very superficial level. While nearly all students visited the website, just over half viewed the information only once. Many students looked at each separate webpage for a very limited time, with half the students viewing at least 6 webpages per minute. The webpages for each OSCE station which focussed on students' proximity to the pass-fail mark were visited much more frequently than those pages that gave a much more detailed breakdown of how the student performed within the stations, or in their performance across stations. Very few students viewed the 'next steps' pages which gave students guidance in how to use the information. Use of the feedback website was not homogenous, with a wide range in the number of webpages viewed and the number of times the website was visited. The students who had only just passed the OSCE were the least engaged; none visited the website more than twice, even though they would potentially have much to gain from feedback. Although the students who passed all stations typically looked at more webpages, their profile of usage was not particularly reassuring; they made more use of aspects of the website which compared their performance with other students, rather than looking at the pages which demonstrated why they have performed strongly. The conclusion from this study was that it is not enough simply to provide feedback and expect students to make full use of the opportunities provided. While this quantitative study provided useful information on how students engage with feedback in a summative context, it failed to provide an explanation of the reasons why learners may, or may not, engage with feedback offered in these circumstances.

The study in **Chapter 3** therefore employed in-depth interviews to explore the individual barriers to feedback receptivity. One of the main findings was the perceived power of the summative assessment culture. The fear of failure and the subsequent "punishment" of needing resits loomed large in students' minds. Learners felt that more attention was focussed on those students who failed, while students who passed felt ignored. This meant that exams were seen as a hurdle, not a stepping stone to future learning. Additionally, students experienced a rollercoaster of strong emotions: fear of taking the exam, relief when it's over, fear of getting the results, intense relief when passed. Interactions with

Summary

others reinforced the need to avoid failure rather than use the feedback information as an aid towards improved clinical performance. Teachers talked about it being a hurdle unrelated to real life as a doctor. Discussions between students focussed on whether they needed to resit or not. Students tended to share the feedback information more with parents, partly to seek approval. Parents reinforced the need to just pass. Students had been used to succeeding in exams at school, but they now had to adjust to not coming top of the class. This led to a sense of futility, which led in turn to an attitude of not trying too hard, and instead focussing on just passing the assessment. There was a disconnection between assessment and future clinical learning. Feedback was seen as relevant for a specific exam or a specific question; it was felt that it might help performance in future exams but was rarely seen as relevant for clinical workplace. Clinical teachers rarely discussed feedback from assessment to help students apply learning in workplace. Overall, students did not appear motivated to make use of feedback.

Although the study in chapter 3 set out to explore the individual barriers to feedback receptivity, it demonstrated that factors within a medical school's assessment culture were influential. The study in **Chapter 4** therefore sought to understand the factors within medical school assessment systems in more detail. Focus groups were held in three different medical schools with very different assessment cultures. Assessment systems either increased or decreased personal agency. Receptivity to feedback was increased if agency was increased. Within assessment, agency was increased by providing elements of choice, enabling students to demonstrate their knowledge to themselves or their tutors. Within feedback, agency was enhanced by enabling students to act independently in accessing feedback. Institutions reduced agency by controlling how feedback was used, which led students to play an 'assessment game'. Students recognised the need for limits to agency. Standardisation of assessments was seen as fair, as long as the assessment had relevance to their work as future doctors. Assessments were often perceived as irrelevant to future clinical work as a doctor, which had two adverse effects. Feedback received in the workplace before the assessment was ignored, as it would not help students to pass the assessment. Secondly, feedback received following the assessment was not seen as relevant for clinical learning. Grades provided reassurance and clarity to students that they were on track. But they did not act as an incentive for students to aim higher – rather the reverse, as high grades seemed unrealistic to obtain. Not having grades caused uncertainty, but was seen as a more authentic preparation for real life as a doctor and allowed a more nuanced view of students' strengths and weaknesses. The uncertainty helped to drive students to aspire towards achieving optimal competence. Mentoring helped students to interpret feedback and could support the fostering of agency if the scaffolding was reduced as time went on through the course. A long-term relationship with a mentor enabled trust to develop. Mentors could see all the feedback relating to the student, which enabled new feedback to be put into context. They could challenge students' inaccurate self-assessments.

These studies demonstrated how assessment cultures have remained rooted in a behaviourist approach to medical education: students are rewarded for passing and punished for failing assessments. Furthermore, the studies supported a move towards a more constructivist approach to assessment, empowering the students to take a more active role. Other constructivist elements were a clear call for more authenticity in assessments and the benefits of mentoring, with gradual descaffolding as the course progresses.

The previous studies had highlighted the need to change the assessment culture in order to enhance feedback receptivity. The study described in **Chapter 5** used participatory design methods to explore how a mixed group of stakeholders might attempt to change the assessment culture to one in which students made more use of post-assessment feedback. There was recognition that current assessments could be inauthentic, so that performance in assessments did not necessarily reflect clinical workplace learning. As a result, inauthentic assessments could reward last-minute learning away from the workplace. There was often a lack of time or opportunity for students to demonstrate their knowledge and skills. There was also recognition that feedback was often ignored if the assessments were inauthentic. However, implementation of more authentic assessment was seen as very challenging, as increasing an assessment's authenticity would lead to a reduction in its reliability. Participants had experienced efforts to improve the authenticity of individual assessments and found these attempts to be often clumsy and unhelpful. Current inauthentic assessments were seen as having some advantages in that they prepared learners for similar assessments early in their postgraduate career.

The provision of feedback without grades was considered but was perceived to be very difficult to implement for a number of reasons. Firstly, it was felt that learners would attempt to convert narrative feedback into a numerical mark or grade even if none was offered. This in turn could be misleading, as there was felt to be a mismatch between current narrative feedback and the numerical mark; students with a mark close to the pass-fail boundary often received quite encouraging feedback, while excellent students often received quite critical feedback with the aim of stretching the learner. Examiners felt that the numerical mark was more important than the narrative feedback. Grades were felt to provide students with clarity and reassurance. Despite awareness of the literature supporting the provision of feedback without grades, participants did not believe this evidence as it did not resonate with their own personal experience.

Mentoring was seen as a beneficial way to encourage students to engage meaningfully with feedback but was also regarded as presenting enormous implementation challenges. While it was recognised that mentoring needed to occur soon after the provision of assessment-related feedback, enough time needed to be allowed to ensure the process was meaningful. Similarly, selection of appropriate mentors provided other practical challenges, with a tension between the need for a long-term relationship with the tutor versus the credibility of feedback coming from a subject specialist.

Summary

There was evidence that personal and collective beliefs were influential in ensuring that participants were reluctant to change. There was a strong shared belief in the primacy of the summative assessment paradigm, which meant that the pass-fail cut score, and the need to exclude unsafe doctors, was seen as the most important purpose of assessments. Allied to this was a need for objectivity and rigour in order for the process to be fair. Participants' prior experiences as learners reinforced their belief in the summative assessment. However, when participants had experienced a difference assessment culture, there was openness to change.

Within the group setting, senior members in particular often expressed their beliefs in a tacit or implicit manner. Beliefs were expressed more explicitly in the context of individual interviews, with the result that resistance to change was not fully apparent from the group setting. Beliefs did not appear easily influenced by the comments of others during the group discussion. Despite dissatisfaction with many aspects of the current assessment culture, and recognition of a need for more meaningful engagement with feedback, there appeared to a shared common assumption that "summative assessment is the way things are done round here" with a result that summative assessment was seen as more important than feedback. This study demonstrated the enormity of the challenge for assessment programme designers who wish to bring about a change in assessment culture.

Finally **Chapter 6** considered the answers to the research questions in the context of the wider research literature. The challenging conclusion was that, although the practical difficulties of delivering feedback in summative assessment context can be overcome, learners will often not be receptive to the feedback provided, as the summative assessment culture is dominant and stifles the learning potential available from assessments. However, when potential change of assessment culture was considered, faculty and students were very reluctant as they maintained strong beliefs in the benefits of summative assessment, often because of their own prior experiences. Despite this, the weight of educational evidence suggests that maintaining the status quo is not an acceptable solution. The thesis concludes by calling for a paradigmatic shift in assessment and feedback culture. When designing a programme of assessment, medicine faces particular challenges because of the need for regulators to reassure the public that their doctors are competent. In order to assist clinician-educators to make the leap of faith towards an assessment culture which fosters learning, analogies are drawn from the field of healthcare. Beliefs in assessments have similarities with patients' beliefs about illness. The thesis argues that these beliefs should be explored in a way that fosters mutual respect, to try to reach compromises which are reasonably acceptable to all stakeholders within an assessment programme. While accepting the significant challenges of implementing long-term mentoring, the thesis argues that there is great potential that clinicians and students would subsequently find the assessment system more rewarding in the end, in much the same way that clinicians and patients find continuity of care extremely beneficial. After considering the strengths and limitations of the thesis, suggestions are made for further research. Although the need for

coherent, well-designed assessment programmes is increasingly accepted, it is essential to research more closely the actual implementation of change in the assessment culture. Suggestions for how this may be achieved are considered.

Samenvatting

Feedback in de context van ‘*high stakes*’-toetsing: kan summatief formatief zijn?

Feedback kan een zeer krachtig middel zijn; een van de meest krachtige manieren om de prestaties te bevorderen. Helaas wordt feedback vaak niet optimaal benut. Dit komt bijvoorbeeld doordat de feedback niet op de juiste wijze wordt gegeven of ontvangen of doordat ontvangers niet weten hoe zij de feedback moeten vertalen naar concrete leeractiviteiten die tot positieve veranderingen leiden. Er is tot op heden weinig aandacht besteed aan de rol van feedback in de context van summatieve toetsing. Het doel van dit proefschrift was dan ook om stil te staan bij de moeilijkheden die bij het ontvangen van feedback in deze context komen kijken, en meer specifiek om de gedeelde opvatting dat toetsen een formatief dan wel een summatief doel hebben en dat het lastig is om beide doelen in één enkele toets te verenigen ter discussie te stellen.

Er zijn verschillende redenen om aan te nemen dat het verstandig zou zijn om de formatieve en summatieve doelen van toetsing met elkaar te verenigen. ‘*High stakes*’-toetsen genereren doorgaans grote hoeveelheden gegevens over de sterke en zwakke punten van een student, maar deze informatie wordt teruggebracht tot een binair voldoende/onvoldoende-besluit over bevordering naar de volgende fase. Hoewel deze aanpak begrijpelijk is, kan dit misleidend zijn, aangezien zo veel waardevolle informatie over de student in het proces verloren gaat; bovendien bereidt een dergelijke aanpak de studenten niet voor op zelfstandig zelfsturend leren. Gedetailleerde informatie uit toetsen kan studenten helpen inzicht te verkrijgen in hun eigen sterktes en aandachtspunten. Praktijktoetsen zijn zeer duur in het medisch onderwijs, dus vanuit financieel oogpunt is het ook zinvol om zoveel mogelijk voordeel uit deze toetsen te behalen.

Aan de andere kant brengt het geven van feedback in deze omstandigheden praktische knelpunten met zich mee, zoals een behoefte aan een verhoogde inzet van staf, hoewel dit deels met behulp van informatietechnologie mogelijk ondervangen kan worden. De ontvankelijkheid van studenten voor feedback vormt echter een groter aandachtspunt; het is niet zeker of zij in deze context op een zinvolle manier met de feedback aan de slag zullen gaan. Wellicht zijn er individuele of cultuurgebonden factoren die de ontvankelijkheid voor feedback belemmeren, maar hier is geen onderzoek naar gedaan in de context van summatieve toetsing. Het is bekend dat summatieve toetsing andere aspecten van het leerproces nadelig beïnvloedt, dus het is aannemelijk dat deze toetswijze ook ongewenste gevolgen kan hebben voor de ontvankelijkheid voor feedback. Hoewel er voorstellen zijn gedaan om de summatieve toetscultuur om te vormen tot een cultuur waarbij de nadruk ligt op toetsen om te leren, ontbreekt er wetenschappelijk bewijs dat aantoont hoe dit zodanig kan worden ingevoerd dat belanghebbenden er ook daadwerkelijk bij betrokken raken.

Dit proefschrift beoogde daarom drie centrale onderzoeksvragen te beantwoorden:

1. Hoe gaan studenten aan de slag met feedback in een ‘*high stakes*’-situatie?
2. Welke individuele en instellingsgebonden factoren beïnvloeden en belemmeren de ontvankelijkheid voor feedback in een dergelijke situatie?

3. Kunnen we de toetscultuur omvormen tot een cultuur waarin leren van feedback bevorderd wordt en voorop staat? Wat houdt ons tegen?

De in **Hoofdstuk 2** beschreven studie beoogde de eerste centrale onderzoeksvraag te beantwoorden door met behulp van kwantitatieve methodes te beoordelen in welke mate geneeskundestudenten aan de slag gingen met feedback na een 'high stakes'-stationstoets. Om de inzet van staf zoveel mogelijk te beperken en om snelle levering van de feedback te faciliteren, werd de feedbackinformatie elektronisch via het web beschikbaar gesteld. De digitale format maakte het niet alleen eenvoudiger om feedback te geven, maar registreerde ook hoe studenten de informatie gebruikten. De gegevens betreffende het gebruik van de website waren gekoppeld aan de resultaten die studenten behaald hadden op de toets en aan hun antwoorden op een vragenlijst over studenteigenschappen met betrekking tot leren. Een aanvullende functionaliteit was dat de website studenten hulp bood bij het opstellen van een actieplan na het bestuderen van de feedback. Uit onze analyse bleek dat de meeste studenten iets deden met de digitale feedback die zij na afloop van een summatieve stationstoets ontvingen, zij het op een erg oppervlakkig niveau. Hoewel nagenoeg alle studenten de website hadden bezocht, bekeek net iets meer dan de helft de informatie slechts één keer. Veel studenten bekeken elke afzonderlijke internetpagina slechts heel kort, en de helft van hen bekeken minstens 6 internetpagina's per minuut. De pagina's die per station aantoonde hoe ver de student van een voldoende verwijderd was werden veel vaker bezocht dan de pagina's die een veel meer gedetailleerde specificatie gaven van hoe de student gepresteerd had binnen de stations en over de diverse stations. Erg weinig studenten bekeken de pagina's met vervolgstappen die studenten op weg hielpen om iets met de informatie te doen. Het gebruik van de feedbackwebsite verschilde erg van student tot student: er was een grote spreiding in het aantal internetpagina's dat zij bekeken en het aantal keer dat zij de website bezochten. De studenten die de stationstoets met een nauwe marge hadden behaald waren het minst betrokken; geen van hen bezocht de website meer dan twee keer, ondanks dat zij mogelijk veel baat zouden hebben van de feedback. Hoewel de studenten die voor alle stations geslaagd waren doorgaans meer internetpagina's bekeken, was hun gebruiksprofiel niet bijzonder geruststellend; in plaats van te kijken naar de pagina's die lieten zien waarom zij zo goed gepresteerd hadden, maakten zij meer gebruik van aspecten van de website die hun prestaties met die van andere studenten vergeleken. De conclusie uit deze studie was dat het niet volstaat om studenten eenvoudigweg van feedback te voorzien en te verwachten dat zij de gegeven kansen volledig benutten. Hoewel deze kwantitatieve studie zinvolle informatie heeft verschaft over hoe studenten in een summatieve context omgaan met feedback, gaf het geen verklaring van de redenen waarom studenten wel of niet van feedback gebruik maken in deze omstandigheden.

De studie in **Hoofdstuk 3** maakte daarom gebruik van diepte-interviews om te onderzoeken welke individuele factoren de ontvankelijkheid voor feedback belemmeren. Een van de belangrijkste bevindingen was de waargenomen invloed van de summatieve toetscultuur.

Angst om te falen en voor de daaruit volgende “straf” te moeten herkansen hing als een donkere wolk boven het hoofd van de studenten. Zij hadden het gevoel dat de aandacht voornamelijk uitging naar de studenten die de toets niet hadden gehaald, terwijl de studenten die waren geslaagd zich gepasseerd voelden. Met andere woorden, de examens werden eerder gezien als een horde dan als opstapje naar toekomstig leren. Bovendien hadden studenten het gevoel zich in een achtbaan van sterke emoties te bevinden: angst voor het afleggen van het examen, opluchting wanneer het voorbij was, angst voor het krijgen van de uitslag, grote opluchting wanneer ze waren geslaagd. Interacties met anderen versterkten de behoefte om falen te voorkomen in plaats van dat de daaruit vloeiende feedbackinformatie werd aangewend als hulpmiddel voor het verbeteren van de klinische prestaties. Docenten spraken van een horde die niets te maken had met het echte leven van een arts. Gesprekken tussen studenten gingen er voornamelijk over of ze wel of niet moesten herkansen. Studenten hadden de neiging om de feedbackinformatie meer met de ouders te delen, deels omdat ze goedkeuring zochten. Ouders versterkten de behoefte om gewoon te slagen. Studenten waren gewend met glans te slagen voor toetsen op school, maar nu moesten ze zich aanpassen aan het feit dat ze niet meer de beste van de klas waren. Als gevolg hiervan voelden zij zich kansloos, wat er weer toe leidde dat studenten zich niet al te veel inspanden en zich in plaats daarvan richtten op het simpelweg behalen van de toets. Er was geen verband tussen toetsing en toekomstig klinisch leren. Feedback werd gezien als relevant voor een specifieke toets of specifieke vraag; men zag het nut ervan in voor het eventueel verbeteren van prestaties op toekomstige toetsen, maar zelden voor de klinische werkplek. Klinische docenten bespraken de toetsfeedback zelden met studenten om hen te helpen het leren op de werkplek toe te passen. Over het algemeen leken de studenten niet gemotiveerd om van de feedback gebruik te maken.

Hoewel het doel van de studie in hoofdstuk 3 was om te onderzoeken welke individuele factoren de ontvankelijkheid voor feedback belemmeren, werd ook aangetoond dat factoren met betrekking tot de toetscultuur van een medische faculteit een belangrijke rol spelen. **Hoofdstuk 4** poogde daarom meer licht te werpen op de factoren binnen toetssystemen van medische faculteiten. Er werden focusgroepen gehouden in drie verschillende medische faculteiten met zeer uiteenlopende toetsculturen. Toetssystemen hadden ofwel een versterkend dan wel een beperkend effect op de persoonlijke invloed. De ontvankelijkheid voor feedback nam toe naarmate er meer verantwoordelijkheid bij de student werd gelegd. Binnen toetsing werd de persoonlijke invloed verhoogd door sommige keuzes aan de student te laten, zodat zij hun kennis aan henzelf of aan hun tutoren konden laten zien. Op het niveau van feedback werd de persoonlijke invloed versterkt door studenten in staat te stellen de feedback op zelfstandige wijze te verkrijgen. Instellingen beperkten de invloed van studenten door te bepalen hoe de feedback werd gebruikt, waardoor de studenten een “toetsspel” gingen spelen. De studenten zagen in dat er grenzen gesteld moesten worden aan hun invloed. Het standaardiseren van toetsen werd goedgekeurd, zolang de toets relevant was voor hun werk als toekomstig arts. Toetsen werden vaak gezien als irrelevant voor het toekomstige klinisch werk als arts, wat twee nadelen met zich meebracht. In de

eerste plaats deden studenten niets met de feedback die zij vóór de toets op de werkplek ontvingen, omdat deze hen niet hielp om de toets te halen. Ten tweede beschouwden zij de feedback die zij na afloop van de toets ontvingen niet als relevant voor klinisch leren. Cijfers stelden studenten gerust en boden inzicht in de vraag of ze wel of niet bij waren. Ze zorgden er echter niet voor dat studenten hun doelen naar boven bijstelden; het tegenovergestelde was eerder het geval, aangezien het behalen van hoge cijfers onrealistisch leek. Geen cijfers zorgde voor onzekerheid, maar werd gezien als een meer authentieke voorbereiding op het echte leven als arts en bood studenten een meer genuanceerd inzicht in hun sterktes en aandachtspunten. Door de onzekerheid werden studenten aangespoord te streven naar optimale competentie. Begeleiding van een mentor hielp studenten de feedback te interpreteren en kon de persoonlijke invloed versterken, mits de ondersteuning in de loop van de opleiding geleidelijk minder werd. Een langdurige relatie met een mentor bood ruimte voor het opbouwen van vertrouwen. Het stelde mentoren in staat alle op de student van toepassing zijnde feedback in te zien, waardoor nieuwe feedback in context kon worden geplaatst. Zij konden de studenten confronteren met hun eventuele foute zelfbeoordelingen.

Deze studies toonden aan hoe toetsculturen nog steeds verankerd zijn in een behavioristische benadering van medisch onderwijs: studenten worden beloond wanneer zij de toetsen halen en gestraft wanneer zij deze niet halen. Verder schraagden de studies een overgang naar een meer constructivistische benadering van toetsing, waarbij studenten in staat worden gesteld een actievere rol te vervullen. Andere constructivistische aspecten waren een duidelijke oproep tot meer authenticiteit in toetsen en de voordelen van begeleiding door een mentor, met geleidelijke afbouw van ondersteuning in de loop van de opleiding.

De voorgaande studies hebben benadrukt dat de toetscultuur eerst moet veranderen wil men de ontvankelijkheid voor feedback vergroten. De in **Hoofdstuk 5** beschreven studie onderzocht met behulp van participatief-ontwerpmethodes hoe een gemengde groep belanghebbenden de toetscultuur zou proberen om te vormen tot een cultuur waarin studenten meer gebruik maken van de feedback die zij na de toets ontvangen. Er werd toegegeven dat de huidige toetsen mogelijk niet authentiek waren, zodat de prestaties op toetsen niet per definitie een weerspiegeling gaven van het leren op de klinische werkplek. Hierdoor kon het zijn dat niet-authentieke toetsen leren op het laatste moment buiten de werkplek beloonden. Studenten hadden vaak te weinig tijd of kansen om hun kennis en vaardigheden te tonen. Ook werd erkend dat de feedback vaak in de wind werd geslagen wanneer toetsen niet authentiek waren. De invoering van meer authentieke toetsen werd echter als iets zeer lastigs gezien, omdat een toename in de authenticiteit gepaard zou gaan met een afname van de betrouwbaarheid van de toets. Deelnemers hadden pogingen meegemaakt om de authenticiteit van afzonderlijke toetsen te verbeteren, maar vonden deze vaak onhandig en nutteloos. Een van de voordelen die zij in de huidige niet-authentieke

toetsen zagen was dat deze de studenten voorbereidden op vergelijkbare toetsen die zij in het begin van hun vervolgopleiding zouden krijgen.

De optie om feedback te geven zonder cijfers werd overwogen, maar dit werd om allerlei redenen gezien als zeer lastig om in te voeren. In de eerste plaats was men de mening toegedaan dat zelfs wanneer er geen punt gegeven werd studenten zouden proberen om de narratieve feedback te vertalen naar een numerieke beoordeling of cijfer. Dit kon misleidend zijn, aangezien men vond dat de huidige narratieve feedback en numerieke beoordeling niet met elkaar overeenstemden. Studenten die met hun cijfer dicht bij de voldoende/onvoldoende-grens zaten ontvingen namelijk vaak redelijk aanmoedigende feedback, terwijl de feedback die uitmuntende studenten ontvingen vaak erg kritisch was met de bedoeling de student nog meer uit te dagen. Examinatoren vonden de numerieke beoordeling belangrijker dan de narratieve feedback. Men vond dat cijfers de studenten duidelijkheid verschafte en geruststelden. Ondanks dat deelnemers ervan op de hoogte waren dat het geven van feedback zonder cijfers door de literatuur werd ondersteund, geloofden zij niet in dit bewijs omdat het niet strookte met hun eigen persoonlijke ervaring.

Begeleiding van een mentor werd gezien als een nuttige manier om studenten aan te moedigen op een zinvolle manier met de feedback om te gaan, maar tegelijkertijd werden grote problemen voorzien met de invoering ervan. Hoewel men inzag dat mentorbegeleiding zo snel mogelijk na het geven van toets-gerelateerde feedback diende plaats te vinden, moest er ook voldoende tijd worden geboden om de zin van het proces te kunnen garanderen. Ook kwamen er andere praktische problemen bij de selectie van geschikte mentoren om de hoek kijken, bijvoorbeeld omdat de behoefte aan een langdurige relatie met de tutor botste met de geloofwaardigheid van feedback afkomstig van een vakspecialist.

Er was bewijs dat persoonlijke en gemeenschappelijke overtuigingen in grote mate garandeerden dat deelnemers van verandering afkerig waren. Er heerste een intense, gedeelde overtuiging dat het summatieve toetsmodel superieur was, wat inhield dat de voldoende/onvoldoende-grensscore, en de behoefte om onveilige arts en artsen uit te sluiten, als belangrijkste doel van de toetsen werd gezien. Hieraan verbonden was een behoefte aan objectiviteit en zorgvuldigheid om een eerlijk proces te kunnen garanderen. Het geloof in de summatieve toets werd versterkt door de ervaringen die de deelnemers zelf hadden gehad als student. Wanneer de deelnemers echter een andere toetscultuur hadden ervaren, stond men open voor verandering.

In groepsverband spraken de oudere leden hun overtuigingen vaak op een stilzwijgende of impliciete manier uit. Tijdens de individuele interviews werden de overtuigingen explicieter uitgesproken, waardoor uit het groepsverband niet helemaal duidelijk werd dat zij van verandering afkerig waren. Overtuigingen bleken niet gemakkelijk beïnvloed te worden door de opmerkingen van anderen tijdens de groepsdiscussie. Ondanks dat er

ontevredenheid bestond over vele aspecten van de huidige toetscultuur en men inzag dat studenten meer op een zinvolle manier aan de slag moesten gaan met de feedback, bleek er een gedeelde opvatting te heersen dat “summatieve toetsing is zoals we het hier aanpakken”, waardoor summatieve toetsing als belangrijker werd beschouwd dan feedback. Deze studie toonde aan hoe immens groot de uitdaging is waarvoor programmaontwikkelaars die een verandering in toetscultuur willen doorvoeren, worden gesteld.

Ten slotte werden in **Hoofdstuk 6** de antwoorden op de onderzoeksvragen beschouwd in de context van de bredere onderzoeksliteratuur. De prikkelende conclusie was dat, ook al is het mogelijk de praktische moeilijkheden die het geven van feedback in een summatieve toetscontext met zich meebrengt te boven te komen, studenten vaak niet open zullen staan voor de feedback, omdat de summatieve toetscultuur overheerst en deze de leerkansen die toetsen te bieden hebben in de kiem smoort. Echter, wanneer een mogelijke verandering in toetscultuur werd overwogen, waren stafleden en studenten erg afkerig omdat zij hevig geloofden in de voordelen van summatieve toetsing, meestal door hun eigen eerdere ervaringen. Desalniettemin weegt het onderwijskundige bewijs te zwaar om te opereren dat het handhaven van de status quo een aanvaardbare oplossing is. Het proefschrift sluit af met een oproep tot een paradigmaverschuiving in toets- en feedbackcultuur. Bij het ontwerpen van een programma van toetsing doen zich in het geval van geneeskunde specifieke moeilijkheden voor omdat er regulerende maatregelen nodig zijn om de mensen te verzekeren dat hun artsen competent zijn. Om arts-opleiders te helpen bij de vertrouwenssprong naar een toetscultuur die het leren bevordert, worden vergelijkingen getrokken met voorbeelden uit de gezondheidszorg. Toetsovertuigingen vertonen gelijkenissen met opvattingen van patiënten over ziekte. Dit proefschrift stelt dat deze overtuigingen onderzocht zouden moeten worden op een manier die wederzijds respect stimuleert, zodat men kan proberen compromissen te sluiten die voor alle belanghebbenden in een toetsprogramma aanvaardbaar zijn. Hoewel het proefschrift onderkent dat de invoering van langdurige begeleiding door een mentor behoorlijk wat moeilijkheden met zich meebrengt, stelt het dat de kans groot is dat klinici en studenten het toetsstelsel uiteindelijk meer zullen gaan waarderen, vergelijkbaar met de manier waarop klinici en patiënten de continuïteit van de zorg uiterst zinvol achten. Na een beschouwing van de sterke en zwakke punten van het proefschrift worden suggesties gedaan voor toekomstig onderzoek. Hoewel het belang van samenhangende, goed ontwikkelde toetsprogramma's steeds meer wordt ingezien, is het essentieel dat de daadwerkelijke invoering van verandering in de toetscultuur nader wordt onderzocht. Er wordt een uiteenzetting gegeven van mogelijke manieren om dit te bereiken.

Valorization

**Feedback in the context
of high-stakes assessment:
Can summative be formative?**

Relevance

The results of this research have significant social and economic implications. The traditional model of summative, end-of-year assessments, followed by resits for those who have not reached the required standard, is very expensive. This is particularly true in the case of the assessment of practical clinical skills, as they involve many senior clinicians, simulated and real patients, administrators, clinical equipment and a suitable location. If such assessments offer predominantly binary pass-fail decisions about results, ignoring the nuances of strengths and weaknesses hidden within the assessments, this represents poor value for money. Furthermore, the traditional model also has implications for society, as all learners who pass a summative assessment at the end of a medical undergraduate course are deemed safe to practise medicine, whatever flaws are hidden within the overall passing mark, as was outlined in Chapter 1.

This thesis sought to explore if these problems could be simply solved by providing cost-effective feedback to all students after summative assessments. A web-based portal provided a low-cost method of providing feedback to all students after summative assessments. As we demonstrated in Chapter 2, almost all students viewed the feedback. However, engagement with the feedback was typically very superficial and brief; this was particularly the case for students who had barely passed the assessment. Thus, although this method of feedback adds little to overall assessment costs, it fails to ensure that learners adequately engage with feedback.

The results of the studies described in Chapters 3 and 4 demonstrated that investing in improving the quantity and quality of feedback is unlikely to have much impact if it is given in the context of high-stakes assessments, as the summative assessment culture is so powerful and dominant that it hinders receptivity to feedback. However, the study in Chapter 4 demonstrated that feedback is valued if it is part of an appropriate assessment culture. Our research sought to clarify the elements of the assessment culture that are important for enhancing feedback receptivity. Some elements, such as mentoring, may present very significant up-front economic costs, but offer the potential for significant long-term savings, by intervening earlier in learners' careers and providing appropriately targeted remediation. Our research also calls for a move away from the traditional behaviourist approach to assessment, in which learners are rewarded for passing and punished for failing, towards a constructivist approach in which learners are much more active participants in the assessment programme. This represents a major change in society's attitude to assessment. The important lesson for cost-effective assessment design is that it is a waste of valuable time and money to simply superimpose a feedback mechanism onto a punitive, behaviourist assessment programme. Even apparently 'cheap' options are expensive if they have little benefit. Extensive redesign of the assessment culture may seem more costly but in the end should turn out to be substantially more cost-effective.

Target groups

The group who have most to potentially gain from this research is the patient community. The ultimate aim of this thesis is to improve patient care by fostering the development of doctors who are receptive to feedback, capable of interpreting it, then acting on it in order to address potential weaknesses in their clinical care. Patient care has undergone a paradigm shift in the last hundred years, from a culture in which patients were typically passive participants to one in which they are active consumers of healthcare, able and willing to challenge clinicians and express their own wishes. At present, patients have little involvement in the assessment of future doctors. It would be helpful for patients to have more insight into assessment processes; they may provide a way to challenge some vested interests within the academic and medical community. Our research has demonstrated that learners are more receptive to feedback when the assessment is more authentic, that is, more representative of real patient care.

Regulators of healthcare, such as the General Medical Council in the UK, should find the research findings interesting, if somewhat challenging. These regulators fulfil important roles in ensuring that entrants to the healthcare profession have reached a satisfactory minimum standard. Much of their focus has been on summative assessments as a way of reliably determining whether this standard has been attained. However, in recent years they have also focussed attention on the performance of existing doctors, ensuring that members of the profession are maintaining appropriate standards and continuing to learn throughout their career. Regulators may be unsettled by the notion that the summative assessments discourage engagement with feedback, especially among academically weaker learners. On the other hand, current programmes of appraisal and revalidation of existing doctors include many of the elements which encourage engagement with feedback. Elements include a significant degree of autonomy and agency and a focus on qualitative judgements rather than a numerical, reductionist pass-fail approach.

Students should also find the research findings relevant. Inherently, students want to become competent clinicians who will thrive in their chosen profession. They have a vested interest in successfully navigating the assessment programme, but also want to maintain their skills as doctors. It will be helpful for learners to understand the flaws in existing assessment programmes in order for them to support the paradigm shift needed in assessment culture. Academically weaker students, who just manage to scrape through programmes dominated by summative assessments, potentially have much to gain from this paradigm shift which would help them to engage with feedback. The study described in Chapter 5 demonstrated, however, that current students' beliefs about assessment culture are strongly influenced by their prior experiences and it cannot be assumed that they will automatically embrace a very different assessment and feedback culture. However, exposure to elements of a different assessment culture should support the drive towards change. Indeed, where assessment programmes have been radically changed in this way,

the students typically become strong proponents of the new culture. Ultimately, students will become the doctors, academics and even regulators of tomorrow, so it is important they understand and embrace the implications of this research in order to bring about the paradigm shift needed.

These results should also be of interest to those who organise assessment in postgraduate medical settings, such as, in the UK, the royal colleges. Their assessment programmes have been gradually moving to incorporate more workplace-based assessment and feedback. Trainees often view such assessments as summative hurdles to navigate, rather than learning opportunities. Such attitudes may at least partly be fostered by these learners' experiences in a summative assessment-dominated undergraduate system. Graduates from redesigned assessment programmes may therefore be better prepared to engage with opportunities in the postgraduate setting. However, royal colleges also need to ensure that they design their assessment programmes carefully in such a way as to support learning.

Although more research in other settings is clearly needed, the findings of this thesis should be of interest to those educating other students, whether in healthcare or elsewhere. Arguably, meaningful engagement with feedback should help in the development of critical thinking skills, which should have benefits for society as a whole.

Activities & Products

Our research findings, and the implications arising from them, should directly influence the redesign of assessment programmes in medical education. We have identified elements which should be included within assessment programmes in order to encourage receptivity to feedback and thereby enhance learning. The findings have already been begun to be incorporated into the author's own medical school, Keele University, UK, as it redesigns its assessment programme.

Many other medical schools are considering transforming their assessment programmes. To assist with this process, the findings have been disseminated through a number of presentations at international conferences, as well as invited talks at several UK medical schools. Chapters 2 – 5 have all been published in peer-reviewed international journals. At the time of writing, the study in Chapter 2 has been cited 29 times since its publication in July 2013. The study described in Chapter 3 has been cited 18 times and downloaded 1400 times since it was first published online in June 2014. In the 9 months since the study in Chapter 4 was published, it has been cited 7 times and downloaded 1400 times. In the two months since Chapter 5's study has been published, the article has been accessed more than 900 times.

Innovation

This research is innovative because there had previously been virtually no empirical work done on receptivity to feedback in the context of summative assessment. The findings clearly demonstrate how the powerful, dominant nature of summative assessments stifles the learning which could occur afterwards. Our research has been the first to uncover the elements within assessment cultures which help or hinder receptivity to feedback, and has clearly demonstrated the need to move away from a purely behaviourist approach to assessment in medical education to one which also incorporates many constructivist elements. This research therefore advances the field in a way which should have significant implications for all programmes which are dominated by summative assessments.

Schedule & Implementation

Dissemination of the research findings has already been taking place. In addition, plans to implement many of the findings are already well underway in the author's own medical school as it redesigns its assessment programme. It is likely to take some time for the ideas to be adopted on a wide scale. It is difficult to estimate costs at present as it is unclear whether there are immediate market opportunities.

Acknowledgements

Although completing the PhD part-time has been a long and at times arduous process, it has been only a small part of my journey through medical education. With apologies for anyone I have inadvertently omitted, I would like to express my deep gratitude to those who have made this PhD possible.

Firstly, I am appreciative of the education I received at Manchester University Medical School. In particular, David Metcalfe stimulated my curiosity in the complexity and uncertainty of general practice; many years later, I have retained my fascination with, and love of, general practice. Tim David stimulated my interest in research by asking a thought-provoking question in a paediatric clinic; the end result was my first published paper.

After I became a GP, Carl Whitehouse gave me my first opportunity to become involved with the teaching of medical students, by allowing me to join the excellent group helping students to communicate effectively with patients. This enabled me to become better acquainted with Tim Dornan, then Hospital Dean in Salford. His enthusiasm was infectious and fostered in me a desire to deliver the best possible experience for students. Carl's retirement led to Val Wass's arrival. Val encouraged me to develop my academic interests by studying for a Masters in Medical Education at Cardiff University with Lynne Allery and her team. I'll never forget the lively and stimulating discussions about educational theory and practice we had there, even if it always seemed to be raining.

All this time, awareness of Maastricht was bubbling in the background. Tim David, Tim Dornan, Carl and others had visited there in order that Manchester could introduce problem-based learning into its curriculum. Tim Dornan and Val were completing PhDs at Maastricht, Cees van der Vleuten had given an inspiring keynote address to the Manchester Medical Education conference, while Lambert Schuwirth came over to Cardiff to run the assessment module with characteristic style. Lambert stimulated my interest in assessment, which soon led to my involvement in the redesign of the OSCE in Manchester. This in turn led to my interest in the potential for feedback after summative assessments.

After obtaining my Masters from Cardiff and moving to Keele to continue my academic involvement, it therefore seemed almost inevitable that I should feed my academic curiosity a little more by enrolling for a PhD at Maastricht. The quality of supervision and (of course) feedback was second to none. Indeed, I learnt a lot about the need to be receptive to feedback as I progressed through the PhD. Some might say I was a little 'high maintenance' in having four supervisors, but they all helped me in different ways. Cees brought incredible clarity of thought and showed how it was possible to say a great deal with relatively few words; his enviable efficiency in email remains an endless source of fascination to me. Lambert always managed to make me look at things from a different perspective; each study benefitted from his sideways view. I'm glad he decided to maintain his involvement when he left for Australia. Karen helped me patiently when I was still getting to grips with SPSS and also showed me the importance of ensuring a seamless

Acknowledgements

transition from one paragraph to another; my writing has improved considerably as a result. Finally, Val was an enormous source of support from beginning to end, encouraging me along at every step, cajoling where necessary but always incredibly warm and understanding. She also taught me how to express ideas concisely but effectively. I'm immensely grateful to all.

The PhD could not have been completed without much support from others behind the scenes. Hilary Harris and Hilary Thompson, my colleagues when I moved to my practice in Brooklands, were very sympathetic to my pursuit of academic interests in addition to clinical commitments. My subsequent colleagues have also kindly allowed me flexibility to attend conferences. At Keele, Kirsty Hartley helped a lot with the recruitment of students, as well as ensuring that the 'day job' of delivering assessments continued smoothly. Adrian Molyneux demonstrated a great flair for innovation as he developed the feedback website and the related app for the iPads. Natalie Cope provided very helpful field notes to support the fourth study. Andy Hassell kindly helped to protect my time at Keele to enable me to maintain momentum with my PhD studies. From Maastricht University, Angelique van den Heuvel efficiently translated my summary into Dutch and Lilian Swaen carefully co-ordinated supervisors' diaries so we could Skype across three countries. I am most grateful for the help from colleagues at Datawyse who have guided me through the process of printing the finished thesis.

Closer to home, the help and support I received from my family was even more valuable. I was told before starting my PhD that I would learn a lot about life; certainly a lot happened in my life during this time, with the death of my mother, my marriage to Ilona and the arrival of our daughter, Scarlett-Maria. Mum loved and supported me all the way, making many sacrifices to assist with my education. She nurtured in me a love of libraries and showed me many different ways to search for information long before the internet era. More importantly, she taught me a great deal about the importance of holding steadfast to my values and principles. I'm forever grateful. She was very patient and understanding even when she felt I was taking on too much work; she will be relieved the PhD is over at last. Dad has followed my progress with great interest, especially the vagaries of the peer review process. He has taught me a lot about the need for careful attention to detail. Sadly he is not well enough to travel to Maastricht, but he will be there in spirit. My brothers, Anthony and David, have also been very supportive. My mother-in-law, Rosmarie, has been very helpful, even allowing me to Skype from her home for supervisions on occasion.

Special mention is reserved for the two very important ladies in my life. Ilona got to know me while I was already studying for the PhD. She has shown the most incredible love, patience and support and always believed it was possible for me to complete the dissertation even when I was starting to doubt myself. The arrival of Scarlett-Maria has brought us such great joy that it is impossible to express with mere words. She has also shown great patience during all the times when Daddy was holed up in his office. It is a

reflection of the dominant nature of a thesis that one of her first words was PhD! Without Ilona and Scarlett-Maria, this dissertation would not have been completed, so it is only fitting that it is dedicated to them. Finally, I will have more time for family, which is infinitely more important than feedback.

Curriculum Vitae

Christopher Harrison was born in 1967. After attending St Ambrose College in Hale Barns, he studied medicine at the University of Manchester, graduating in 1990. After training in general practice, he was awarded Membership (with distinction) of the Royal College of General Practitioners in 1994. After eight years as a partner at Barlow Medical Centre in Didsbury, Manchester, he moved in 2004 to Brooklands Medical Practice, also in Manchester, where he remains as partner. He was awarded Fellowship of the Royal College of General Practitioners in 2006.

His involvement with medical education began in 1999, when he was appointed as a part-time lecturer in the University of Manchester, with responsibility for teaching students patient-centred communication. After gradually increasing his time commitment, he studied for a Masters in Medical Education at Cardiff University, graduating with distinction in 2008. While studying for his Masters, he developed an interest in assessment and was responsible for radically redesigning the OSCE in Manchester. In 2009 he was appointed as the Academic Lead for Assessment at Manchester Medical School.

In 2010 he moved to Keele University School of Medicine as Senior Lecturer in Medical Education with a remit to improve feedback. He led the introduction of paperless marking of OSCEs with the provision of verbal feedback via an iPad app. In 2013 he was appointed Director of Assessment. He is currently redesigning the assessment programme for the medical school's new curriculum, due to be implemented from 2018.